STIC Search Report
Biotech-Chem Library

Scientific & Technical Information Center
USPTO

STIC Database Tracking Number: 117719

TO: Celine Qian
Location: rem/2a89/2c70
Art Unit: 1636
Thursday, April 01, 2004

Case Serial Number: 09/913878

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Qian,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

STIC SEARCH RESUL FEEDBACK FORM

USPTO

Biotech-Chem Librarv

Questions about the scope or the results of the search? Contact ***the searcher or contact***:

Mary Hale, Information Branch Supervisor
571-272-2507 Remsen E01 D86

Voluntary Results Feedback Form

- *I am an examiner in Workgroup:* ☐ *Example: 1610*
- *Relevant prior art **found**, search results used as follows:*
  - ☐ 102 rejection
  - ☐ 103 rejection
  - ☐ Cited as being of interest.
  - ☐ Helped examiner better understand the invention.
  - ☐ Helped examiner better understand the state of the art in their technology.

  Types of relevant prior art found:
  - ☐ Foreign Patent(s)
  - ☐ Non-Patent Literature
    (journal articles, conference proceedings, new product announcements etc.)
- *Relevant prior art **not found:***
  - ☐ Results verified the lack of relevant prior art (helped determine patentability).
  - ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.

sira
Search and Information Resources Administration

Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.
Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rnpm** and **.rnpn**
Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapm** and **.rapn**

Because they contain data that is confidential, the results of Pending database searches should not be left in the case .

ATTN: Ed Hart

Access DB# 117719

SEARCH REQUEST FORM

CRFE

Scientific and Technical Information Center

Requester's Full Name Celine Qian Examiner #: 78710 Date: 3/25/04
Art Unit 1636 Phone Number 306-0283 Serial Number 09/713878
Mail Box and Bldg Room Location 2A87 Results Format Preferred (circle) PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need. mej

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Isolation & characterization of an Crassa silencing gene

Inventors (please provide full names): Giuseppe, Hanino

Earliest Priority Filing Date: 11/6/2000

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search ✓1) SEQ ID NO:1 NA-8045
✓2) nucleic acid encoding seq ID NO:2 AA-1402
✓3) aa 710 - aa 1282 of seq ID NO:2
✓4) nucleic acid encoding 710-1282 of seq ID NO:2

mej

BEST AVAILABLE COPY

STAFF USE ONLY		Type of Search		Vendors and cost where applicable	
Searcher		NA Sequence (#)	3	STN	reverse to NA
Searcher Phone #		AA Sequence (#)	2+1	Dialog	
Searcher Location		Structure (#)		Questel/Orbit	
Date Searcher Picked Up	3/29/04	Bibliographic		Dr. Link	
Date Completed		Litigation		Lexis/Nexis	
Searcher Prep & Review Time		Fulltext		Sequence Systems	064/05P
Clerical Prep Time		Patent Family		WWW/Internet	
Online Time		Other		Other (specify)	

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         March 31, 2004, 02:06:52 ; Search time 3597.57 Seconds
                                   (without alignments)
                                   6903.427 Million cell updates/sec

Title:          US-09-913-878A-2_COPY_710_1282
Perfect score:  3010
Sequence:       1 GVGRMSRSVAKRIRDVLGLG..........EKWCAITPEAMDKSGANYDS 573

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:      6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09913878/runat_29032004_082008_8366/app_query.fasta_1.2254
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878_@CGN_1_1_7684_@runat_29032004_082008_8366 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3010	100.0	4206	8	NCR133528	AJ133528 Neurospor
	2	3010	100.0	8045	6	AX034425	AX034425 Sequence
	3	275.5	9.2	3505	8	NTA011576	AJ011576 Nicotiana
	4	268.5	8.9	3807	8	AY148431	AY148431 Arabidops
c	5	268.5	8.9	82584	8	NCB13B3	BX284762 Neurospor
	6	266.5	8.9	1538	8	ATH011977	AJ011977 Arabidops
	7	265.5	8.8	6688	8	AF468822	AF468822 Diaporthe
	8	263	8.7	5992	3	DDI314910	AJ314910 Dictyoste
	9	262	8.7	3731	6	AR145904	AR145904 Sequence
	10	262	8.7	3731	8	LERDRP	Y10403 L.esculentu
	11	260.5	8.7	5731	3	DDI314909	AJ314909 Dictyoste
c	12	259	8.6	165701	8	AP004357	AP004357 Oryza sat
	13	258	8.6	3591	6	AX204840	AX204840 Sequence
	14	254	8.4	3524	8	AF443073	AF443073 Phomopsis
c	15	254	8.4	132716	2	AC105931	AC105931 Magnaport
	16	250	8.3	44658	8	SPAC6F12	Z98533 S.pombe chr
	17	247	8.2	6847	8	AY029284	AY029284 Neurospor
	18	247	8.2	7178	8	AF500110	AF500110 Neurospor
	19	244	8.1	6987	8	AY049072	AY049072 Diaporthe
	20	235.5	7.8	6204	8	AF411016	AF411016 Neurospor
	21	230	7.6	29132	3	CEF10B5	Z48334 Caenorhabdi
	22	227.5	7.6	121337	8	OSJN00104	AL606653 Oryza sat
c	23	225.5	7.5	132699	8	AC006917	AC006917 Genomic s
	24	225	7.5	5025	3	AF159144	AF159144 Caenorhab
	25	221	7.3	4978	3	AF117611	AF117611 Dictyoste
	26	220.5	7.3	4013	8	AF239718	AF239718 Arabidops
	27	220.5	7.3	6863	6	AX204839	AX204839 Sequence
c	28	220.5	7.3	80367	8	AC012329	AC012329 Arabidops
c	29	220.5	7.3	104204	8	ATT9C5	AL132964 Arabidops
	30	219	7.3	5208	3	AF159143	AF159143 Caenorhab
	31	215.5	7.2	2639	8	AK101440	AK101440 Oryza sat
	32	215.5	7.2	5280	8	AF268093	AF268093 Arabidops
c	33	211.5	7.0	74406	8	F2P3	AF080120 Arabidops
	34	211.5	7.0	108598	8	ATT22B4	AL049876 Arabidops
	35	211.5	7.0	198301	8	ATCHRIV31	AL161531 Arabidops
	36	190.5	6.3	4300	3	AF293414	AF293414 Giardia i
	37	189.5	6.3	6103	3	AF393743	AF393743 Giardia i
c	38	180.5	6.0	31521	3	CEY47H10A	AL132848 Caenorhab
	39	178	5.9	2132	8	AK064209	AK064209 Oryza sat
	40	177	5.9	1618	8	AK064379	AK064379 Oryza sat
c	41	173.5	5.8	40786	3	AF537961	AF537961 Branchios
	42	164.5	5.5	6204	8	AF411018	AF411018 Neurospor
	43	161.5	5.4	4579	8	PHY011979	AJ011979 Petunia h
c	44	160	5.3	38019	3	CEF26A3	Z78419 Caenorhabdi
	45	151.5	5.0	80554	2	AP004143	AP004143 Oryza sat

ALIGNMENTS

RESULT 1

```
NCR133528
LOCUS       NCR133528               4206 bp    DNA     linear   PLN 10-MAY-1999
DEFINITION  Neurospora crassa qde-1 gene, partial.
ACCESSION   AJ133528
VERSION     AJ133528.1  GI:4803726
KEYWORDS    qde-1 gene; RNA-dependent RNA polymerase.
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Cogoni,C. and Macino,G.
  TITLE     Gene silencing in neurospora crassa requires a protein homologous
            to RNA-dependent RNA polymerase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4206)
  AUTHORS   Cogoni,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-1999) Cogoni C., Dept. Biotecnologie Cellulari ed
            Ematologia,, Universita di Roma La Sapienza, Viale Regina Elena
            324, 00161 Rome, ITALY
FEATURES             Location/Qualifiers
     source          1..4206
                     /organism="Neurospora crassa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5141"
     gene            1..4206
                     /gene="qde-1"
     CDS             1..>4206
                     /gene="qde-1"
                     /function="gene silencing"
                     /note="putative"
                     /codon_start=1
                     /product="RNA-dependent RNA polymerase"
                     /protein_id="CAB42634.1"
                     /db_xref="GI:4803727"
                     /db_xref="GOA:Q9Y7G6"
                     /db_xref="SPTREMBL:Q9Y7G6"
                     /translation="MNPITPRKRNSPVEEIINRLNNDYNLGLQCVADTTLTPHRRKEL
                     AESDEDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLP
                     WSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNT
                     KRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQC
                     TKRAKGKLSDNVAAAAAPPVPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQ
                     GTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKS
                     HIDNQPLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAP
                     VVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDV
                     TDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTA
                     PLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQ
                     WLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFRP
                     DVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKTYAI
                     MTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFG
                     SAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQL
                     LPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRV
                     SHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRV
                     GRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQ
                     RVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEM
                     PLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNY
                     KERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDP
                     MYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWDPDLA
                     SYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQVY
                     EKWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHKSPKF
                     VWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGDGSEYP
                     DPEVYEVLGDDDFDGIGFTGNGDY"
ORIGIN

Alignment Scores:
Pred. No.:              2.16e-246       Length:         4206
Score:                  3010.00         Matches:        573
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0
```

```
US-09-913-878A-2_COPY_710_1282 (1-573) x NCR133528 (1-4206)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2128 GGTGTAGGCCGCATGTCGCGAAGCGTGGCCAAGAGGATACGCGATGTTCTCGGTTTGGGT 2187

Qy         21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2188 GATGTGCCCTCTGCTGTGCAAGGGCGGTTTGGTTCGGCCAAGGGAATGTGGGTTATCGAC 2247

Qy         41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2248 GTTGACGACACAGGCGATGAGGATTGGATCGAGACATACCCGTCCCAGCGCAAGTGGGAA 2307

Qy         61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2308 TGCGACTTCGTTGATAAACATCAACGTACCCTCGAAGTCCGGAGCGTGGCTTCTGAACTG 2367

Qy         81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2368 AAGTCAGCTGGTCTCAACCTACAGCTGTTACCTGTCCTGGAAGATAGAGCCAGGGACAAG 2427

Qy        101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2428 GTGAAGATGCGCCAGGCAATCGGTGACCGTCTTATCAACGATTTGCAACGACAGTTCAGC 2487

Qy        121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2488 GAGCAAAAGCATGCTTTGAATCGCCCAGTGGAATTTCGCCAATGGGTTTACGAGAGTTAT 2547

Qy        141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2548 TCCAGTCGCGCAACTCGAGTCAGCCACGGCCGTGTGCCTTTTCTTGCTGGGCTACCTGAC 2607

Qy        161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2608 AGTCAAGAGGAGACACTGAACTTCTTGATGAACAGTGGGTTCGATCCCAAGAAGCAAAAG 2667

Qy        181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2668 TACTTGCAAGACATCGCCTGGGATCTTCAAAAGCGGAAATGTGACACGTTGAAGTCCAAG 2727

Qy        201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2728 CTGAACATCCGTGTCGGTCGATCAGCATACATTTACATGATTGCCGATTTCTGGGGTGTG 2787

Qy        221 LeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGluSer 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2788 CTTGAGGAAAATGAGGTTCATGTCGGATTCTCCTCAAAGTTCAGAGACGAGGAGGAGTCT 2847

Qy        241 PheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2848 TTTACACTCCTATCGGACTGTGATGTCCTCGTGGCGCGATCCCCAGCCCATTTCCCTAGT 2907

Qy        261 AspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIle 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2908 GATATCCAACGGGTTCGAGCAGTCTTCAAGCCAGAGCTCCACAGTCTCAAGGATGTAATC 2967

Qy        281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2968 ATCTTCTCTACTAAAGGAGATGTACCGCTTGCTAAGAAGCTATCTGGTGGAGACTACGAC 3027

Qy        301 GlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGlu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3028 GGCGATATGGCCTGGGTCTGCTGGGATCCGGAGATCGTCGATGGTTTCGTCAATGCGGAA 3087

Qy        321 MetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGln 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3088 ATGCCTCTGGAGCCCGACCTGTCTAGGTACCTAAAGAAGGACAAAACGACTTTCAAACAA 3147

Qy        341 LeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIle 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3148 CTTATGGCCTCACACGGCACGGGCTCAGCGGCCAAAGAGCAGACTACATACGATATGATC 3207
```

```
Qy       361 GlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3208 CAGAAGAGCTTCCATTTCGCCCTGCAGCCCAACTTCTTGGGCATGTGCACTAACTACAAA 3267

Qy       381 GluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSer 400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3268 GAAAGGCTCTGTTACATCAACAATAGTGTGTCTAACAAGCCGGCCATCATTCTTAGTTCA 3327

Qy       401 LeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrp 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3328 CTGGTGGGAAACCTCGTCGATCAGAGCAAGCAAGGTATTGTCTTTAACGAAGCAAGCTGG 3387

Qy       421 AlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLys 440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3388 GCTCAATTGCGTAGGGAACTGCTTGGCGGTGCATTGTCCCTTCCTGACCCAATGTACAAG 3447

Qy       441 SerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSer 460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3448 AGCGACAGTTGGCTCGGGCGCGGAGAGCCTACCCACATTATTGACTACCTGAAATTCTCC 3507

Qy       461 IleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAla 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3508 ATCGCCAGGCCTGCGATTGACAAGGAACTGGAAGCCTTCCACAATGCCATGAAAGCGGCC 3567

Qy       481 LysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhe 500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3568 AAGGATACAGAAGACGGCGCTCACTTTTGGGATCCGGATCTCGCTTCCTACTACACGTTC 3627

Qy       501 PheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsn 520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3628 TTCAAGGAGATTAGCGACAAGTCGCGATCGTCCGCACTGCTATTCACGACTCTGAAGAAC 3687

Qy       521 ArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAsp 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3688 CGTATCGGCGAAGTCGAGAAAGAATATGGCAGGTTGGTCAAAAACAAGGAGATGAGAGAC 3747

Qy       541 SerLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThr 560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3748 AGCAAGGACCCCTACCCTGTCCGCGTCAACCAGGTTTATGAAAAATGGTGCGCCATCACG 3807

Qy       561 ProGluAlaMetAspLysSerGlyAlaAsnTyrAspSer 573
             |||||||||||||||||||||||||||||||||||||||
Db      3808 CCTGAGGCGATGGACAAATCCGGAGCAAATTATGATTCT 3846
```

```
RESULT 2
AX034425
LOCUS       AX034425                8045 bp    DNA     linear   PAT 22-SEP-2000
DEFINITION  Sequence 1 from Patent WO0050581.
ACCESSION   AX034425
VERSION     AX034425.1  GI:10303130
KEYWORDS    .
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Cogoni,C. and Macino,G.
  TITLE     Isolation and characterization of a n. Crassa silencing gene and
            uses thereof
  JOURNAL   Patent: WO 0050581-A 1 31-AUG-2000;
            UNIV ROMA (IT) ; COGONI CARLO (IT) ; MACINO GIUSEPPE (IT)
FEATURES             Location/Qualifiers
     source          1. .8045
                     /organism="Neurospora crassa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5141"
     CDS             2447. .6655
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC10121.1"
                     /db_xref="GI:10303131"
                     /db_xref="REMTREMBL:CAC10121"
                     /translation="MNPITPRKRNSPVEEIINRLNNDYNLGLQCVADTTLTPHRRKEL
                     AESDEDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLP
                     WSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNT
                     KRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQC
                     TKRAKGKLSDNVAAAAAPPVPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQ
                     GTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKS
                     HIDNQPLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAP
                     VVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDV
                     TDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTA
                     PLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQ
                     WLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFRP
                     DVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKTYAI
                     MTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFG
                     SAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQL
                     LPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRV
                     SHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRV
                     GRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQ
                     RVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEM
                     PLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNY
                     KERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDP
                     MYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWDPDLA
                     SYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQVY
                     EKWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHKSPKF
                     VWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGDGSEYP
                     DPEVYEVLGDDDFDGIGFTGNGDY"
ORIGIN

Alignment Scores:
Pred. No.:              5.09e-246       Length:          8045
Score:                  3010.00         Matches:         573
Percent Similarity:     100.00%         Conservative:    0
Best Local Similarity:  100.00%         Mismatches:      0
Query Match:            100.00%         Indels:          0
DB:                     6               Gaps:            0

US-09-913-878A-2_COPY_710_1282 (1-573) x AX034425 (1-8045)

Qy         1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4574 GGTGTAGGCCGCATGTCGCGAAGCGTGGCCAAGAGGATACGCGATGTTCTCGGTTTGGGT 4633

Qy        21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4634 GATGTGCCCTCTGCTGTGCAAGGGCGGTTTGGTTCGGCCAAGGGAATGTGGGTTATCGAC 4693

Qy        41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4694 GTTGACGACACAGGCGATGAGGATTGGATCGAGACATACCCGTCCCAGCGCAAGTGGGAA 4753

Qy        61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4754 TGCGACTTCGTTGATAAACATCAACGTACCCTCGAAGTCCGGAGCGTGGCTTCTGAACTG 4813

Qy        81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4814 AAGTCAGCTGGTCTCAACCTACAGCTGTTACCTGTCCTGGAAGATAGAGCCAGGGACAAG 4873

Qy       101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4874 GTGAAGATGCGCCAGGCAATCGGTGACCGTCTTATCAACGATTTGCAACGACAGTTCAGC 4933

Qy       121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4934 GAGCAAAAGCATGCTTTGAATCGCCCAGTGGAATTTCGCCAATGGGTTTACGAGAGTTAT 4993

Qy       141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4994 TCCAGTCGCGCAACTCGAGTCAGCCACGGCCGTGTGCCTTTTCTTGCTGGGCTACCTGAC 5053

Qy       161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5054 AGTCAAGAGGAGACACTGAACTTCTTGATGAACAGTGGGTTCGATCCCAAGAAGCAAAAG 5113
```

```
Qy        181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5114 TACTTGCAAGACATCGCCTGGGATCTTCAAAAGCGGAAATGTGACACGTTGAAGTCCAAG 5173

Qy        201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5174 CTGAACATCCGTGTCGGTCGATCAGCATACATTTACATGATTGCCGATTTCTGGGGTGTG 5233

Qy        221 LeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGluSer 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5234 CTTGAGGAAAATGAGGTTCATGTCGGATTCTCCTCAAAGTTCAGAGACGAGGAGGAGTCT 5293

Qy        241 PheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5294 TTTACACTCCTATCGGACTGTGATGTCCTCGTGGCGCGATCCCCAGCCCATTTCCCTAGT 5353

Qy        261 AspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIle 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5354 GATATCCAACGGGTTCGAGCAGTCTTCAAGCCAGAGCTCCACAGTCTCAAGGATGTAATC 5413

Qy        281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5414 ATCTTCTCTACTAAAGGAGATGTACCGCTTGCTAAGAAGCTATCTGGTGGAGACTACGAC 5473

Qy        301 GlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGlu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5474 GGCGATATGGCCTGGGTCTGCTGGGATCCGGAGATCGTCGATGGTTTCGTCAATGCGGAA 5533

Qy        321 MetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGln 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5534 ATGCCTCTGGAGCCCGACCTGTCTAGGTACCTAAAGAAGGACAAAACGACTTTCAAACAA 5593

Qy        341 LeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIle 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5594 CTTATGGCCTCACACGGCACGGGCTCAGCGGCCAAAGAGCAGACTACATACGATATGATC 5653

Qy        361 GlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5654 CAGAAGAGCTTCCATTTCGCCCTGCAGCCCAACTTCTTGGGCATGTGCACTAACTACAAA 5713

Qy        381 GluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSer 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5714 GAAAGGCTCTGTTACATCAACAATAGTGTGTCTAACAAGCCGGCCATCATTCTTAGTTCA 5773

Qy        401 LeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrp 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5774 CTGGTGGGAAACCTCGTCGATCAGAGCAAGCAAGGTATTGTCTTTAACGAAGCAAGCTGG 5833

Qy        421 AlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLys 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5834 GCTCAATTGCGTAGGGAACTGCTTGGCGGTGCATTGTCCCTTCCTGACCCAATGTACAAG 5893

Qy        441 SerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSer 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5894 AGCGACAGTTGGCTCGGGCGCGGAGAGCCTACCCACATTATTGACTACCTGAAATTCTCC 5953

Qy        461 IleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAla 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5954 ATCGCCAGGCCTGCGATTGACAAGGAACTGGAAGCCTTCCACAATGCCATGAAAGCGGCC 6013

Qy        481 LysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhe 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6014 AAGGATACAGAAGACGGCGCTCACTTTTGGGATCCGGATCTCGCTTCCTACTACACGTTC 6073

Qy        501 PheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsn 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6074 TTCAAGGAGATTAGCGACAAGTCGCGATCGTCCGCACTGCTATTCACGACTCTGAAGAAC 6133

Qy        521 ArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAsp 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6134 CGTATCGGCGAAGTCGAGAAAGAATATGGCAGGTTGGTCAAAAACAAGGAGATGAGAGAC 6193

Qy        541 SerLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThr 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6194 AGCAAGGACCCCTACCCTGTCCGCGTCAACCAGGTTTATGAAAAATGGTGCGCCATCACG 6253

Qy        561 ProGluAlaMetAspLysSerGlyAlaAsnTyrAspSer 573
              |||||||||||||||||||||||||||||||||||||||
Db       6254 CCTGAGGCGATGGACAAATCCGGAGCAAATTATGATTCT 6292
```

```
RESULT 3
NTA011576
LOCUS       NTA011576               3505 bp    mRNA    linear   PLN 25-FEB-2003
DEFINITION  Nicotiana tabacum mRNA for RNA-directed RNA polymerase.
ACCESSION   AJ011576
VERSION     AJ011576.1  GI:4138281
KEYWORDS    dRP gene; RNA-directed RNA polymerase.
SOURCE      Nicotiana tabacum (common tobacco)
  ORGANISM  Nicotiana tabacum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Nicotiana.
REFERENCE   1
  AUTHORS   Schiebel,W., Pelissier,T., Riedel,L., Thalmeir,S., Schiebel,R.,
            Kempe,D., Lottspeich,F., Saenger,H.L. and Wassenegger,M.
  TITLE     Isolation of an RNA-directed RNA polymerase-specific cDNA clone
            from Tomato
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3505)
  AUTHORS   Wassenegger,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1998) Wassenegger M., Max-Planck-Institut of
            Biochemistry Viroidresearch, Am Klopferspitz 18A, 82152
            Martinsried, GERMANY
FEATURES             Location/Qualifiers
     source          1..3505
                     /organism="Nicotiana tabacum"
                     /mol_type="mRNA"
                     /cultivar="Petit Havana SRI"
                     /db_xref="taxon:4097"
                     /tissue_type="leaves"
     gene            1..3505
                     /gene="RdRP"
     CDS             20..3370
                     /gene="RdRP"
                     /codon_start=1
                     /evidence=experimental
                     /product="RNA-directed RNA polymerase"
                     /protein_id="CAA09697.1"
                     /db_xref="GI:4138282"
                     /db_xref="GOA:Q9ZRY7"
                     /db_xref="SPTREMBL:Q9ZRY7"
                     /translation="MGKTIQVFGFPYLLSAEAVKKFLENHTGNGTVYALEVKQSKGGR
                     RAFAKVQFANNKSAEFIIDLASKGLYYGPSYLKAWEMKTDIVQPRTYVEHMDGVTLNF
                     GCQISDKKFAVLGSKDVSIKFGIGLKKIYFFLSHASADYKLQLSYENIWQVVLHRPYG
                     QNAQFLLLQLFGAPRIYKRLEDSCYSFFKETPDDQWVRTTDFTPSWIGLSSSLCLEFR
                     NGVQLPNFSESFFYYKESMNQFILQTGFTFSFSQKLALVPIVHLPEGIELPYKILFKI
                     SSLIQHGCFPGLALNFNFFQLVDPRRRNIACIEHALEKLYYLKECCYDPVRWLTEQYD
                     EYLKGRQLPKSPPITLDDGLVYVRRVVVTPCKVYFCGPEVNVSNRVLRNYSEDINNFL
                     RVSFVDEEWEKIHSTDLLPRASTGNGTRTDIYERILSTLRNGFIIGDKRFEFLAFSSS
                     QLRDNSVWMFASRPGLTANDIRTWMGDFRQIRNVAKYAARLGQSFGSSRETLSVGRHE
                     VEVIPDVACSLHGTNYIFSDGIGKISADFAHRVTIKCGLQYTPSSFQIRYGGYKGVVA
                     VDPYSSMKLSLRKSMLKYESNNIKLDVLGWSKYQPCYLNRQLVTLLSTLGVIDDVLEQ
                     KQNEAVDQLDAILHDSLKAQEALELMSPGENTNILKEMLNCGYMPDAEPFLSMMLQTF
                     RASKLLDLRTRTRIFIPNGRTMMGCLDESRTLEYGQVFVQFSGAGRRQFFEESHPFND
                     SGSANCDFILKGNVVVAKNPCLHPGDIRVLRAVDVPALHHMVDCVVFPQKGKRPHPNE
                     CSGSDLDGDIYFVCWDPDLIPPRQVQSMDYTPAPTTQLDHDVTIEEVEEYFTNYIIND
                     SLGIIANAHVVFADREPDMAMSDPCKQLAQLFSIAVDFPKTGVPAEIPSQLRPKEYPD
                     FMEKPDKPTYPSEKVIGKLFQKVKNKTPQASSIATFTRDVARKSYDSDMIVDGFEDYI
                     DEAFYYKSEYDNKLGNLMDYYGIKTEAEILSGGIMKASKTFDRRKDAEAIGVAVRCLR
                     KEARAWFKRRSDIDDMLAKASAWYHVTYHHTYWGLYNEGLKRDHFISFPWCVYDQLIQ
                     IKKAKARKRPVPHLSSLGSQLSRKLVIW"
ORIGIN

Alignment Scores:
Pred. No.:            1.7e-13        Length:          3505
```

```
Score:                    275.50     Matches:         123
Percent Similarity:       35.65%     Conservative:    82
Best Local Similarity:    21.39%     Mismatches:      215
Query Match:              9.15%      Indels:          155
DB:                       8          Gaps:            16

US-09-913-878A-2_COPY_710_1282 (1-573) x NTA011576 (1-3505)

Qy        1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
Db     1604 GGAATTGGAAAAATATCTGCTGACTTTGCTCATAGAGTTACCATAAAATGTGGCCTTCAA 1663
Qy       21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
Db     1664 TATACTCCATCTTCTTTTCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGCTGTTGAT 1723
Qy       41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
Db     1724 --------------------------------CCATATTCATCAATGAAG--------- 1741
Qy       61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
Db     1742 ------------------------------TTATCTTTGAGAAAGAGCATGTTGAAATAT 1771
Qy       81 LysSerAlaGlyLeuAsnLeuGlnLeuLeu------------------ProValLeuGlu 94
Db     1772 GAATCAAACAACATAAAGTTAGATGTCCTGGGATGGAGCAAATATCAGCCTTGTTATCTT 1831
Qy       95 AspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAsp 114
Db     1832 AATCGTCAACTGGTTACTCTCTTGTCTACTCTTGGGGTGATAGATGATGTTTTAGAGCAG 1891
Qy      115 LeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGln 134
Db     1892 AAGCAAAATGAAGCAGTAGATCAGCTTGATGCTATTTTGCATGATTCTTTGAAGGCACAG 1951
Qy      135 TrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPhe 154
Db     1952 GAAGCTTTAGAATTGATGTCT--------------------------------------- 1972
Qy      155 LeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPhe 174
Db     1973 -----------CCTGGAGAGAATACTAATATTCTCAAGGAGATGCTAAATTGTGGTTAT 2020
Qy      175 AspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCys 194
Db     2021 ATGCCTGATGCTGAGCCGTTTCTTTCAATGATGCTGCAAACTTTCCGGGCTTCGAAGTTG 2080
Qy      195 AspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIle 214
Db     2081 CTCGATTTGCGGACTAGAACAAGAATATTTATTCCAAATGGTAGAACGATGATGGGATGC 2140
Qy      215 AlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSer------ 232
Db     2141 TTGGATGAATCCAGAACCCTGGAATATGGTCAGGTGTTTGTTCAGTTTTCTGGTGCTGGA 2200
Qy      233 ------LysPheArgAspGluGluGluSerPhe------ThrLeuLeuSerAspCysAsp 248
Db     2201 CGTAGACAGTTTTTTGAGGAATCGCATCCTTTTAATGACAGTGGATCCGCCAACTGTGAT 2260
Qy      249 ------------------ValLeuValAlaArgSerProAlaHisPheProSerAspIle 262
Db     2261 TTCATTCTCAAGGGAAACGTGGTCGTTGCAAAAAATCCATGCTTGCATCCTGGTGATATT 2320
Qy      263 GlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePhe 282
Db     2321 CGTGTTTTAAGGGCTGTTGATGTTCCAGCTTTGCACCACATGGTTGATTGTGTTGTATTT 2380
Qy      283 SerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAsp 302
Db     2381 CCCCAGAAAGGAAAAAGACCTCATCCAAATGAATGTTCTGGAAGTGATTTGGATGGGGAT 2440
Qy      303 MetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetPro 322
Db     2441 ATCTACTTTGTTTGTTGGGATCCAGACTTGATTCCGCCAAGGCAAGTCCAGTCGATGGAT 2500
Qy      323 LeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMet 342
Db     2501 TATACTCCAGCACCCACCACACAGTTGGATCATGAT------------------------ 2536
Qy      343 AlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLys 362
Db     2537 -----------------------------------GTCACAATTGAGGAAGTTGAAGAG 2560
Qy      363 SerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArg 382
Db     2561 TACTTC------------------------------------ACCAACTAT--------- 2575
Qy      383 LeuCysTyrIleAsnAsnSerValSer--------------------------------- 391
Db     2576 ------ATCATTAATGACAGTTTGGGAATCATAGCAAATGCGCACGTTGTATTTGCAGAC 2629
Qy      392 --------------------AsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
Db     2630 AGAGAACCTGATATGGCCATGAGCGATCCGTGCAAACAACTTGCTCAGCTCTTTTCAATT 2689
Qy      405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
Db     2690 GCAGTTGACTTTCCAAAGACTGGCGTTCCAGCTGAAATACCATCT------CAGTTGCGC 2743
Qy      425 ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAsp---ProMetTyrLysSerAspSer 443
Db     2744 CCTAAAGAATACCCAGACTTCATGGAAAAGCCAGACAAGCCAACCTATCCCTCGGAAAAA 2803
Qy      444 TrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArg 463
Db     2804 GTTATTGGAAAG------------------------------------------------ 2815
Qy      464 ProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThr 483
Db     2816 -----------------------------------CTTTTCCAGAAAGTGAAGAACAAA 2839
Qy      484 GluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGlu 503
Db     2840 ---------------------ACACCCCAGGCTAGCTCCATCGCGACCTTCACAAGGGAT 2878
Qy      504 IleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGly 523
Db     2879 GTTGCCAGGAAATCATATGATAGTGATATGATAGTCGATGGATTCGAAGATTACATTGAC 2938
Qy      524 GluVal---------GluLysGluTyrGlyArgLeuValLysAsn 535
Db     2939 GAAGCTTTTACTACAAAAGTGAATACGACAACAAGCTTGGTAAT 2983

RESULT 4
AY148431
LOCUS       AY148431                3807 bp    mRNA    linear   PLN 12-NOV-2002
DEFINITION  Arabidopsis thaliana RNA-dependent RNA polymerase 1 mRNA, complete
            cds.
ACCESSION   AY148431
VERSION     AY148431.1  GI:24935274
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 3807)
  AUTHORS   Yu,D., Fan,B., MacFarlane,S.A. and Chen,Z.
  TITLE     Analysis of the involvement of an inducible Arabidopsis
            RNA-dependent RNA polymerase in antiviral defense
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3807)
  AUTHORS   Yu,D., Fan,B., MacFarlane,S.A. and Chen,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-SEP-2002) Botany and Plant Pathology, Purdue
```

```
            University, West Lafayette, IN 47907, USA
FEATURES             Location/Qualifiers
     source          1..3807
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
     CDS             128..3451
                     /note="RdRP1"
                     /codon_start=1
                     /product="RNA-dependent RNA polymerase 1"
                     /protein_id="AAN64409.1"
                     /db_xref="GI:24935275"
                     /translation="MGKTIQVFGFPNGVSAEEVKKFLERLTGSGTVYAIKVRQPKKGG
                     PRVYAIVQFTSERHTRLIITAAAERLYYGRSYLKAFEVEQDIVPKPRASLHTISGLKM
                     FFGCQVSTKKFLTLWSAQDVCVSFGIGMRKLHFSFSWYQKDYRLELSYENIWQIDLHS
                     PQGRSSKFLVIQVIGAPKIFEKEDQPINLLFGIMDFYSDGSDEQWIRTTDFTSSSCIG
                     QSTAFCLELPVHLNVPDFRENFANYAEHRASSFLIESGSSYSSNANTLVPVVDPPPGF
                     SLPFEILFKLNTLVQNACLSGPALDLDFYRLLNQKKYDRALIDHCLEKLFHLGECCYE
                     PAHWLRDEYKKWISKGKLPLSPTISLDDGLVYMYRVQVTPARVYFSGPEVNVSNRVLR
                     HYSKYINNFLRVSFVDEDLEKVRSMDLSPRSSTQRRTKLYDRIYSVLRDGIVIGDKKF
                     EFLAFSSSQLRENSAWMFAPIDRITAAHIRAWMGDFDHIRNVAKYAARLGQSFSSSRE
                     TLNVRSDEIEVIPDVEIISLGTRYVFSDGIGKISAEFARKVARKCGLTEFSPSAFQIR
                     YGGYKGVVAVDPNSSKKLSLRKSMSKFESENTKLDVLAWSKYQPCYMNRQLITLLSTL
                     GVTDSVFEKKQREVVDRLDAILTHPLEAHEALGLMAPGENTNILKALILCGYKPDAEP
                     FLSMMLQNFRASKLLELRTKTRIFISGGRSMMGCLDETRTLEYGQVVVQYSDPMRPGR
                     RFIITGPVVVAKNPCLHSGDVRVLQAVNVPALNHMVDCVVFPQKGLRPHPNECSGSDL
                     DGDIYFVCWDQELVPPRTSEPMDYTPEPTQILDHDVTIEEVEEYFANYIVNDSLGIIA
                     NAHTAFADKEPLKAFSDPCIELAKKFSTAVDFPKTGVAAVIPQHLYVKEYPDFMEKPD
                     KPTYESKNVIGKLFREVKERAPPLISIKSFTLDVASKSYDKDMEVDGFEEYVDEAFYQ
                     KANYDFKLGNLMDYYGIKTEAEILSGGIMRMSKSFTKRRDAESIGRAVRALRKETLSL
                     FNASEEEENESAKASAWYHVTYHSSYWGLYNEGLNRDHFLSFAWCVYDKLVRIKKTNL
                     GRRQRQETLERLDHVLRFG"
ORIGIN

Alignment Scores:
Pred. No.:              7.47e-13        Length:         3807
Score:                  268.50          Matches:        137
Percent Similarity:     37.90%          Conservative:   90
Best Local Similarity:  22.87%          Mismatches:     225
Query Match:            8.92%           Indels:         148
DB:                     8               Gaps:           22

US-09-913-878A-2_COPY_710_1282 (1-573) x AY148431 (1-3807)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
                |||:::|||::::::|||            |||:::::::::          |||||
Db       1736 GGAATAGGAAAGATATCAGCTGAATTTGCTAGAAAGGTAGCGAGGAAATGTGGTCTTACA 1795

Qy         21 AspVal---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMet-TrpValIl 39
              :::      |||||||||  |||   |||:::|||      |||||||::: |||::::
Db       1796 GAGTTTTCTCCATCTGCTTTTCAAATCCGTTATGGCGGTTATAAAGGAGTGGTGGCTGTT 1855

Qy         39 eAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTr 59
               :
Db       1856 G------------------------------------------------------------ 1856

Qy         59 pGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGl 79
                                :::         ||||||||  |||   :::|||      |||::
Db       1857 -------------ATCCAAACTCATCAAAGAAA-CTGTCTCTGAGGAAGAGTATGAGCAA 1902

Qy         79 uLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeu------------------ProValLe 93
               :    :::|||        |||    :::|||                    |||
Db       1903 ATTCGAATCGGAGAACACCAAGCTTGATGTTCTGGCGTGGAGCAAGTACCAACCTTGTTA 1962

Qy         93 uGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAs 113
                  :::|||           :::               :::   |||       :::
Db       1963 TATGAACAGACAACTGATTACACTTTTGTCTACTCTTGGAGTTACGGACAGTGTGTTTGA 2022

Qy        113 nAspLeuGlnArgGlnPheSerGluGlnLysHisAla---LeuAsnArgProValGluPh 132
                    ||||||:::       ::::::      |||    |||     |||:::|||
Db       2023 GAAGAAACAAAGGAAGTTGTGGATCGTCTCGACGCAATCTTGACTCATCCTTTGGAA-- 2080

Qy        132 eArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgVa 152
                                                                :::|||   ::
Db       2081 ----------------------------------------------GCTCACGAGGCTCT 2094

Qy        152 lProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSe 172
              :     :::|||             |||                  |||       |||:::
Db       2095 TGGTTTAATGGCT------CCAGGGGAAAACACAAATATTCTCAAGGCATTGATCTTGTG 2148

Qy        172 rGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysAr 192
                |||:::    |||         :::   :::|||         :::       :::   :::
Db       2149 TGGCTATAAACCTGACGCTGAACCTTTCCTTTCAATGATGCTTCAGAATTTCAGGGCATC 2208

Qy        192 gLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTy 212
                |||          |||::::::|||          |||   :::                :::
Db       2209 TAAGTTGTTGGAACTACGGACAAAAACAAGGATTTTCATTTCTGGTGGAAGATCTATGAT 2268

Qy        212 rMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSe 232
                        |||               |||||         :::|||   |||    :::|||
Db       2269 GGGATGCCTAGACGAGACCAGAACGCTGGAGTATGGTCAGGTAGTAGTGCAGTATTCAGA 2328

Qy        232 rLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAl 252
                    |||                                 :::           |||:::||||
Db       2329 TCCCATGAGGCCGGGAAGGCGA---------TTCATCATCACCGGACCTGTTGTTGTTGC 2379

Qy        252 aArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGl 272
              |::::::|||                   |||::::::   ::::::|||||          |||
Db       2380 CAAAAACCCATGCCTGCATTCTGGTGACGTGCGTGTTCTTCAAGCTGTCAATGTCCCAGC 2439

Qy        272 uLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLy 292
              |||:::    :::     |||      ::::::|||       |||||          |||
Db       2440 TTTAAATCATATGGTGGACTGCGTTGTGTTCCCGCAAAAGGGCTTGAGGCCACACCCAAA 2499

Qy        292 sLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIl 312
                :::     |||||   |||   |||||||||:::   :::|||||||||||   |||::
Db       2500 TGAATGTTCTGGGAGTGATTTAGATGGAGATATTTACTTTGTATGTTGGGATCAAGAATT 2559

Qy        312 eValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLy 332
              :|||             :::         |||         |||:::   ::::::   |||
Db       2560 GGTTCCGCCAAGAACGTCTGAACCAATGGACTACACTCCTGAACCAACTCAAATCTTGGA 2619

Qy        332 sLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLy 352
                   |||
Db       2620 TCATGAT----------------------------------------------------- 2626

Qy        352 sGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPh 372
                          |||    :::    :::::::::     |||              :::      :::
Db       2627 -------GTCACAATTGAGGAAGTTGAAGAGTACTTTGCGAACTACATTGTGAATGATAG 2679

Qy        372 eLeuGlyMetCysThrAsnTyr------------------LysGluArgLeuCysTyrIl 386
                |||||:::        |||                          |||||       |||
Db       2680 TTTAGGGATCATCGCGAATGCTCATACCGCCTTTGCGGATAAGGAACCACTC-------- 2731

Qy        386 eAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVa 406
                                :::    |||    |||    |||:::                   ||
Db       2732 -------AAGGCGTTTAGTGACCCATGCATTGAGCTTGCAAAGAAGTTTTCAACTGCCGT 2784

Qy        406 lAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGl 426
              |||         |||    |||:::                         |||    :::   :::
Db       2785 AGACTTTCCAAAAACTGGTGTTGCA------------------GCTGTGATACCGCAACA 2826

Qy        426 uLeu------------LeuGlyGlyAlaLeuSerLeuProAsp---ProMetTyrLysSe 441
                |||                            :::    |||||   |||    |||:::||
Db       2827 TCTTTACGTGAAAGAGTATCCTGATTTCATGGAGAAACCGGATAAGCCAACATATGAGTC 2886

Qy        441 rAspSerTrpLeuGly------------------------ArgGlyGluProThrHisIl 453
              |     :::    :::|||                          |||       |||
Db       2887 GAAGAACGTGATTGGTAAGCTTTTTAGAGAGGTAAAAGAGCGAGCTCCACCGTTGATCTC 2946

Qy        453 eIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeu-------- 470
                |||               :::|||    :::    |||||:::::::
Db       2947 GATCAAATCATTCACGCTTGATGTGGCCTCAAAGTCTTATGATAAAGACATGGAAGTTGA 3006

Qy        471 ----------------------GluAlaPheHisAsnAlaMetLysAlaAlaLysAspTh 483
                                     |||||||||:::         |||||          |||
```

```
Db       3007 CGGATTTGAGGAGTATGTTGATGAAGCTTTCTAC------CAGAAGGCGAATTATGATTT 3060

Qy        483 rGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGl 503
               :::   |||                  :::|||      ||||||            ||
Db       3061 CAAGTTAGGT------------------AATTTGATGGATTACTATGGGATTAAGACTGA 3102

Qy        503 u---------------------IleSerAspLysSerArgSerSerAlaLeuLeuPheTh 516
              |                     |||         |||:::|||            |||||
Db       3103 GGCTGAGATTCTTAGTGGTGGTATAATGAGGATGTCAAAGTCA------------TTCAC 3150

Qy        516 rThrLeuLysAsnArgIleGlyGlu---Val-GluLysGluTyrGlyArg---------- 531
              |      |||   |||      |||   ::: |||:::   :::||||||
Db       3151 C------AAGAGACGAGATGCAGAATCGATTGGAAGAGCGGTTCGGGCGTTGAGGAAAGA 3204

Qy        532 --------------------LeuValLysAsnLysGluMetArgAspSerLys 542
                                  |||:::|||   |||:::|||         :::
Db       3205 AACTTTGTCGTTGTTCAATGCTTCTGAAGAAGAAAATGAGTCAGCAAAGG 3257

RESULT 5
NCB13B3/c
LOCUS       NCB13B3                 82584 bp    DNA     linear   PLN 06-MAR-2003
DEFINITION  Neurospora crassa DNA linkage group II BAC contig B13B3.
ACCESSION   BX284762
VERSION     BX284762.1  GI:28881246
KEYWORDS    .
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Schulte,U., Aign,V., Hoheisel,J., Brandt,P., Fartmann,B.,
            Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 82584)
  AUTHORS   German Neurospora genome,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2003) MIPS, Institut fuer Bioinformatik,
            GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,
            Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:
            G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute
            of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,
            E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT     this contig is an assembly of BAC 13B3 from 1 to: 11950 and BAC
            11C11 from 11951 to: 82584, strain OR74A, BAC clones are available
            at the Fungal Genetic Stock Center http://www.fgsc.net
            Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
            http://www.mwgdna.com
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of linkage groups II
            and V can be viewed at: http://mips.gsf.de/proj/neurospora.
FEATURES             Location/Qualifiers
     source          1. .82584
                     /organism="Neurospora crassa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5141"
                     /chromosome="6"
     repeat_region   3959. .3982
                     /note="24 bp tcatcatcatca tandem repeat"
     gene            5038. .6880
                     /gene="B13B3.010"
     CDS             join(5038. .5145,5216. .6880)
                     /gene="B13B3.010"
                     /note="similarity to clone mgc:8691, similar to amplified
                     in osteosarcoma, homo sapiens, TREMBL:BC000532_1"
                     /codon_start=1
                     /product="conserved hypothetical protein"
                     /protein_id="CAD70506.1"
                     /db_xref="GI:28881247"
                     /translation="MRRPSLALLALSSLPFGSARQPASFSIHQDLLAHPQFEVIFSDS
                     YVFEADAFALLEAANKTPKPTPASDGAHDGSTTRTDLTSAIRESATANADTDNGDESI
                     GGTSPLRETYELIAHPPMRYLCSIPIIAPPPALNKTATELAKAEEAREVTRAYNKGWE
                     LMRGLENQCLHFVSGWWSYQYCYGKSIVQYHAVPNPKGGPPLRDKNSQEYILGTSLPP
                     SSHSQKGKQIEVPNNEQKQLSPPPNTELQAKDNQRYLVQRLDGGTICDLTGRPRTIEI
                     QYHCNPALSGDRIGWIKEVTTCAYLMVIHTPRLCADVAFLPPKETKAHPITCRQIITS
                     DEEALSFNQRRKNTIDSAAAAAAAVTTEDQKQGSESGSPEKLSYQGLTVAGIPIGARR
                     ILPSTHVLPLPRHLQQQRQEAQQGNLLEALTKAAFKADVFGDYGDDNNNNNNHHPKA
                     GKGRKAAGAGKGQSGQKEMKKMRISERDIDKLGLDQQTLDALREEIRAAGLDPDRNLN
                     DEREAGGEIVWEFYADVSGDEDDGAVAEEGKEVFVWYEDEDEGGEPAEAGKDQKESKK
                     GGNGEGSGSGSEEGSKEEYYRDEL"
     exon            5038. .5145
                     /gene="B13B3.010"
                     /number=1
     intron          5146. .5215
                     /gene="B13B3.010"
                     /number=1
     exon            5216. .6880
                     /gene="B13B3.010"
                     /number=2
     repeat_region   7534. .7562
                     /note="29 bp poly(a) tandem repeat"
     gene            12653. .15824
                     /gene="B13B3.020"
     CDS             join(12653. .12719,12910. .13083,13261. .13438,13497. .14132,
                     14195. .15824)
                     /gene="B13B3.020"
                     /note="similarity to transcription activator, Aspergillus
                     oryzae, PIR:JC7252
                     Contains Fungal Zn(2)-Cys(6) binuclear cluster domain
                     [VCDKCRLRRIRCDYQSPCQPCIDAKLQC]"
                     /codon_start=1
                     /product="related to transcription activator"
                     /protein_id="CAD70507.1"
                     /db_xref="GI:28881248"
                     /translation="MMFMWMGAHGRRRGNLAILPGLQLRGSSSPEAQIVRLKTEPGTS
                     TKQSPRPSSSSNATSAAGSSNRPVQPPKVCDKCRLRRIRCDYQSPCQPCIDAKLQCTR
                     NHVPRKRGPRPGRGRLINRLREREAGKKEEQRPSSSGDPEVSDGARSNNTDIYPQNAS
                     VSSTWASDVSTAPETTGLWANEHIFPSRGAFFYLIPRCVELFYQHLYPVMPVLYMPEI
                     QRMDPRVCDLPEKNLLFSLSALTCFRMSGHSLGAEDSPEFWDQAGRFLLNDCLDVRKQ
                     YDLYDNISLSTVVSSIFLASSFFETNQSTKAWVYLREALTFAQELGLEDESTYAGLSP
                     EEALCRQRVFWLLYVMERSFAILRNKPLMLRRAPKLPMTLHAYESRGIHTGFLQLLGI
                     YIPLRDSIIEAWTYGSNSTVDVNTYLALQNQLARPAGSSGGLSIPHPFFPSSPNAAVS
                     SIQMPVSQSCSFSDLSLSSSSISSSPEDPVAPVVEPTASQTAELLVTQQWLRLIIWLS
                     SLRQGYLSWAAENESMHFALPLTIARQTGLVLRSLMQIGGLEATGVAALEVNGMGIFE
                     KIFEIGAWCMNVLDSYDKASLEGQFLTLPNFGRRHSAASGKNKQAARSATSSPADQVM
                     DYEDGEDVDLLDVFMRALSATPTSRKQFAEPLYMWATTRPGGMRIGPGSVLHLDTGGG
                     TGAFDMAQTQPGLAEQIPGLGLNQGLGLDHASGIGAAQGHDASSMMHQQQQPGIWNAA
                     SAYAHGQAPLPGSATGSGAGLSRPDGPGPGGGQGQGMEDVVMGDQASHQGQGFNIDPR
                     LGAGSNVPRGLQHATFGRRPLRPNAPHLPSPLPGQFGGPPTSLYDVPGWSPIDPGGLT
                     RARSRTMGMGQLRGLMAMVQGGPGTPGSSGPSVSQAGS"
     exon            12653. .12719
                     /gene="B13B3.020"
                     /number=1
     intron          12720. .12909
                     /gene="B13B3.020"
                     /number=1
     exon            12910. .13083
                     /gene="B13B3.020"
                     /number=2
     intron          13084. .13260
                     /gene="B13B3.020"
                     /number=2
     exon            13261. .13438
                     /gene="B13B3.020"
                     /number=3
     intron          13439. .13496
                     /gene="B13B3.020"
                     /number=3
     exon            13497. .14132
                     /gene="B13B3.020"
                     /number=4
     intron          14133. .14194
                     /gene="B13B3.020"
                     /number=4
     exon            14195. .15824
                     /gene="B13B3.020"
                     /number=5
```

```
gene            16797. .17552
                /gene="B13B3.030"
gene            complement(join(16797. .16985,17226. .17552))
                /gene="B13B3.030"
CDS             complement(join(16797. .16985,17226. .17552))
                /gene="B13B3.030"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAD70508.1"
                /db_xref="GI:28881249"
                /translation="MHFIKSLLSLGLLSLLRYHAALALPTISDDTINDDTTINDDTLI
                PRVMIPKDPNMNFDPNNKPNVTKYSITDNPDHDVTLICPSPRALLSSGPPPEPGRFST
                TLGRWTTVSWTYQNAPTYKTLAEQAQMVLDQEKCRTTDATNTRPLVAGEAVYGHHDAW
                SVVLKGVVDCT"
exon            complement(16797. .16985)
                /gene="B13B3.030"
                /number=1
intron          complement(16986. .17225)
                /gene="B13B3.030"
                /number=1
repeat_region   17029. .17053
                /note="25 bp poly(t) tandem repeat"
exon            complement(17226. .17552)
                /gene="B13B3.030"
                /number=2
gene            18460. .24617
                /gene="B13B3.040"
gene            complement(join(18460. .24111,24165. .24617))
                /gene="B13B3.040"
CDS             complement(join(18460. .24111,24165. .24617))
                /gene="B13B3.040"
                /note="similarity to am-toxin synthetase (amt), alternaria
                alternata, TREMBL:AF184074_1
                Contains Putative AMP-binding domain signature
                [IQFTSGSTGKPK];ATP/GTP-binding site motif A (P-loop)
                [AAWSKGKT];Cysteine proteases inhibitors signature
                [QQVVQGALSFLGLN];Phosphopantetheine attachment site
                [AVGGDSLRAMRLIAAA]"
                /codon_start=1
                /product="related to AM-toxin synthetase (AMT)"
                /protein_id="CAD70509.1"
                /db_xref="GI:28881250"
                /translation="MQQVVQGALSFLGLNGQSSNATSHLFDTLGLETVSRPLSLPSQK
                NISLLDASLLAGLLLLQRERGEDDRVENFNWGRRTSDGAETSSQFSVKAVLGSEGLDQ
                TASVKNVLEVLKKVTGQDVTEENRTLFFNDGKEAPEKIKEEGYETAQRPNWKFQLELL
                ASEDGEQHDLTLQAYWFKNVPDALTSSQVTIRLNAFTELLGLIFSSPETTPISSLLAP
                LPCDLNQIWAFNAVVPPLIDRTMQDIISEQAAANPSKPAVSSWDGEFTYAEVESKSDL
                LARHLVSRGVKLGGVVPLMFEKSRWTIVALLAVMKAGSAFALTDPTSQPEGRLRVMVE
                QTGGDIIVASASQTELAEVCLPSSSGGSQGQVITVSDDLFTTLSQTPELTSVTLPVIP
                TQTTPLYIQFTSGSTGKPKGVVISHANYTSGALPRAEAVGYKSSSRCFEFASYAFDVS
                IDCMLCTLAVGGTICIPSDADRMNDLGGAILKSGANMAHMTPSVARVLDPEVIKGLDV
                LGLGGEAVSAADAAAWSKGKTSVIVAYGPSECTVGCTVNNVFARDEEAKKKGFTTGNI
                GTGVGGCGWVVDPNDHDRLVPVGAVGELLMEGPVVGLGYLGEEEKTKEVFIEDPKWLV
                AGHGDAHPGRKGRLYKTGDLVRYDADGSGAFVFVGRKDAQVKLRGQRVELVEIEHHLR
                AKLPSGVKIAAEVIKPSGGDPTLVAFLAETSSTGISAKDALQSVVEGEVSFSPELQAA
                LDGIEEVLGIDLPRYMVPNAYIPLREMPSLVSGKIDRKTLRAMGAAMTREQVSGSAKK
                ARQGGEGGAPTTEMEIALHKVWKQLLGTEQEIAAGDSFFAVGGDSLRAMRLIAAARAE
                GVKLTVADVFRYPVLRDMAVVATKVDTQDAGGKDNNEIAPFSLLESDWAIEDVKAEAS
                KNCNIDKDALEDVYPCTPLQEALMALSAKVKEAYVAQRVLKMEGQAAADKLKAAFEAI
                NADCAILRTRIIHSPRGLMQVVVKEPLHWHSATTLAEYLETDRNEEMDLGKPLVRYGL
                IQDGDAYHIVLTMHHALYDGWSMPLVVDRVNQAYQGLVPRKPAAEFKHFINYLNNTLD
                RAACDAYWKDQLAGATGVQFPALPFEGYQTRADSLWELYVSLDGRKPPSSPNTTITLA
                TIVRAAWALVASQYTAGNNDIVFGETLTGRNAPIVGADEIEGPMITTVPVRVCIDREQ
                TVAEYLQQVAEQMITQIPYEHAGLQHIRRLSDDALQACELRTGIVLHPAAGEVGVTEN
                TPANGLVPAGDSEAAQEALKFNTYALMLVCSLEADGFLVMASFDSKTVEMDVVKKALD
                QLQVVVRQLCEDGCGEKKVGQLECLTEKDHEEVKKLAGVKNLTAESDPFVGGVTAEDV
                EGVWIVDAADTARLVPVCAAGELIVETSKELDAPVTLLAETPAWYEGNKKLYKTGQLA
                KFKVDGDSASLCILKASEPKTEVAKKKVASSVTPSAVAATSAKQKTLRKIWAKLLKVK
                EDSIYLGDSFFNRGGDSITAMKLVSEARQQGLKISVAQVFANRTLFDMANVMQTTGTV
                VDASEKVETEYKPFSLLEENNKASIMEVVKHSLADKSWTIADILPTRPLQEIAVRGTV
                ELPRFSLRYELMHFNGQVDKAALFRACHEVVARNEILRTVFVRDASNTCYGAVIDNLI
                TSIAEYEIDDHADVKDFAAQVCRLDSQTRMPYGSSFVKWFFITNGTSQCSLAFRLSHA
                QYDEICLPIMLHQFHQLYLNSPESSIPPSYPFSHFVAHTLRDSIPAAIPYWRNLLSGS
```

```
     SGITRFRPSTPITSRKHFAIHRAVDISSRPRDVTIATLPSAAWALALARHVKSPDVLF

Alignment Scores:
Pred. No.:              4.37e-11        Length:         82584
Score:                  268.50          Matches:        123
Percent Similarity:     37.74%          Conservative:   57
Best Local Similarity:  25.79%          Mismatches:     168
Query Match:            8.92%           Indels:         129
DB:                     8               Gaps:           20

US-09-913-878A-2_COPY_710_1282 (1-573) x NCB13B3 (1-82584)

Qy        1 GlyValGlyArgMetSerArgSerValAlaLysArgIle---ArgAspValLeuGlyLeu 19
            ||||||||  :::|||:::         :::    |||  ||:::::::
Db    42932 GGCGTTGGGATAATTTCGCAGGGTGCCCTCGAGGTGATTCATCGCGAGATTCCCGAATCC 42873

Qy       20 GlyAspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIle 39
                     |||:::  :::|||  |||:::      |||||||||||        :::
Db    42872 AAGGGCTATCCCAATTGCTTACAGGTTCGATGGGCTGGAGCAAAGGGCATGTTGGCTCTC 42813

Qy       40 AspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrp 59
            |||      |||||   :::       |||         |||    |||
Db    42812 GATGCCCGGCTCACTGGGAGGCAAATCTGTATCAGG------GATTCCATGGAGAAG--- 42762

Qy       60 GluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGlu 79
                     |||     :::  :::      ||||||:::      :::||||||:::
Db    42761 ---------TTCCGCAGCCGTGATGAGGAACATCTCGAAATATGCGACATGGCTTCCAAG 42711

Qy       80 LeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAsp 99
                           ||||||  |||::::::    :::||||||||
Db    42710 CCAATTCCTCTGATGCTGAACCGTCAGATGATCAAAATTCTCGAGGAC------------ 42663

Qy      100 LysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPhe 119

Db    42663 ------------------------------------------------------------ 42663

Qy      120 SerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSer 139
                              :::         |||         ||||||      |||
Db    42662 -----------------ATGCGCGCCCCTGCC---------CAGTGGTTCTTGGAGCTT 42630

Qy      140 TyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuPro 159
                              |||                           |||     :::
Db    42629 CAAGAAAAGGAACTCCAACGT-----------------------CTGCGGGCAATCACG 42594

Qy      160 AspSerGlnGluGluThrLeuAsnPheLeu------------------------------ 169
            |||:::   :::         ||||||
Db    42593 GATAACGTCCAGAACGTGGCAACCTTTCTCAAGCTGCAGTGTGTCGGAGATTCTGTACAC 42534

Qy      170 ------------------------------MetAsnSerGlyPheAspProLysLysGln 179
                                          ||||||      |||    ::::::
Db    42533 CTCAGCCAGTTTCTCAAAGACCTTGACAAAATGAAC------ATCGACTACCGGAGAGAC 42480

Qy      180 LysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSer 199
            ::::::|||:::   |||       :::     |||:::          ||||||
Db    42479 CAGTTTCTTCGCGGTATTGTCGAAGCTGTCGTGCTTAGAGAGCTGCGCTTGCTCAAGCAC 42420

Qy      200 LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219
            |||      |||   |||           ::::::  :::    |||      |||
Db    42419 AAGGCTAGAATCCCGGTTCCCTATGGCGTTACTCTGTTTGGTGTCATGGACGAGACAGGC 42360

Qy      220 ValLeuGluGluAsnGluValHisValGlyPheSerSer---------LysPheArgAsp 236
            :::|||   |||   ||||||:::|||    |||    :::       :::|||:::|||
Db    42359 CTATTGCGCGAAGGCGAAGTATACGTGACTTTTGAGACCGTGGACGGTCGCTTCAAGGAC 42300

Qy      237 GluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAla 256
                                          |||:::|||   |||||||||||
Db    42299 CCTCCCACTGCT-----------------GGCCCCGTTGTGGTGACCAGGTCTCCAGCC 42258
                        :::
Qy      257 HisPheProSerAspIleGlnArgValArgAlaValPheLysPro---------GluLeu 273
                |||   ||||||||               |||        |||
Db    42257 TTGCACCCTGGAGACATCCAGATTGCCCACAATGCCATTCCACCGGCAGGGCACCCTCTC 42198

Qy      274 HisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLys 293
```

```
                       |||||:::    |||:::|||||      |||:::       |||       :::
Db      42197 AGAGAACTGAAAAACTGCATCGTGTTCAGTCAAAACGGTGAACGTGACCTGCCTAGTCAG 42138

Qy        294 LeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleVal 313
              ||||||||||||||     ||||||||         |||   ||||||      ||||||
Db      42137 CTGTCGGGCGGTGACTTGGATGGCGACACTTTCAACGTGATCTGGGATCAAAGCATCGTT 42078

Qy        314 ---------AspGlyPheValAsnAlaGluMetPro---LeuGluPro---AspLeuSer 328
                       |||      |||:::   |||   :::||||||       |||:::
Db      42077 GCCACGCTCAGAACTTTCGCGGCGGCCGATTATCCACGTGTCGAGCCCTTGAAACTGAAC 42018

Qy        329 ArgTyrLeuLys---LysAspLysThrThrPhe------LysGlnLeuMetAlaSerHis 345
              |||   :::::    |||||          |||      :::   :::   |||   |||
Db      42017 CGCGAGGTCGAGTCAAAGGACATGGCTGACTTCTTCGTGGAGTTTATGAAAGCCGATCAC 41958

Qy        346 GlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHis 365
                          |||   :::             |||
Db      41957 CTGGGGGTAATTGCGGTTCGGCACATGATCCTCGCGGAC--------------------- 41919

Qy        366 PheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyr 385
                                                          ||||||
Db      41918 ------------------------------------------GAACGC--------- 41913

Qy        386 IleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeu 405
                 |||          :::         :::   |||::::::|||
Db      41912 ---AATGAGGGAACGCTGGATGCAGACTGTCTCAAGCTGGCAGCATTGCACTCTAAAGCG 41856

Qy        406 ValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArg 425
              ||||||   ||||||   ||||||                       :::|||   |||
Db      41855 GTTGACTTCTCCAAATCGGGGATT---------CACGTGGATATCACCGAGTTACCAAGG 41805

Qy        426 GluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp 442
                                                 |||||||||:::   |||
Db      41804 -----------------------------CCACCAATGTATCGGCCTGAT 41784
```

RESULT 6
ATH011977

```
LOCUS       ATH011977            1538 bp    DNA     linear   PLN 12-MAR-2003
DEFINITION  Arabidopsis thaliana partial RdRP gene.
ACCESSION   AJ011977
VERSION     AJ011977.1  GI:4127462
KEYWORDS    RdRP gene; RNA-directed RNA polymerase.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Schiebel,W., Pelissier,T., Riedel,L., Thalmeir,S., Schiebel,R.,
            Kempe,D., Lottspeich,F., Saenger,H.L. and Wassenegger,M.
  TITLE     Isolation of an RNA-directed RNA polymerase-specific cDNA clone
            from Tomato
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1538)
  AUTHORS   Wassenegger,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1998) Wassenegger M., Max-Planck-Institut of
            Biochemistry Viroidresearch, Am Klopferspitz 18A, 82152
            Martinsried, GERMANY
FEATURES             Location/Qualifiers
     source          1..1538
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /tissue_type="leaves"
                     /note="ecotype: Columbia"
     gene            1..1538
                     /gene="RdRP"
     CDS             <1..>1538
                     /gene="RdRP"
                     /codon_start=1
                     /evidence=experimental
                     /product="RNA-directed RNA polymerase"
                     /protein_id="CAA09894.1"
                     /db_xref="GI:4375829"
                     /db_xref="GOA:Q9ZRX5"
                     /db_xref="SPTREMBL:Q9ZRX5"
                     /translation="IRAWMGDFDHIRNVAKYAARLGQSFSSSRETLNVRSDEIEVIPD
                     VEIISLGTRYVFSDGIGKISAEFARKVARKCGLTEFSPSAFQIRYGGYKGVVAVDPNS
                     SKKLSLRKSMSKFESENTKLDVLAWSKYQPCYMNRQLITLLSTLGVTDSVFEKKQREV
                     VDRLDAILTHPLEAHEALGLMAPGENTNILKALILCGYKPDAEPFLSMMLQNFRASKL
                     LELRTKTRIFISGGRSMMGCLDETRTLEYGQVVVQYSDPMRPGRRFIITGPVVVAKNP
                     CLHSGDVRVLQAVNVPALNHMVDCVVFPQKGLRPHPNECSGSDLDGDIYFVCWDQELV
                     PPRTSEPMDYTPEPTQILDHDVTIEEVEEYFANYIVNDSLGIIANAHTAFADKEPLKA
                     FSDPCIELAKKFSTAVDFPKTGVAAVIPQHLYVKEYPDFMEKPDKPTYESKNVIGKLF
                     REVKERAPPLISIKSFTLDVASKSYDKDMEVDGFEEYVDEAFYQKANYDFKLGNLMDY
                     YGIKT"
ORIGIN

Alignment Scores:
Pred. No.:              3.34e-13        Length:          1538
Score:                  266.50          Matches:         121
Percent Similarity:     38.06%          Conservative:    83
Best Local Similarity:  22.57%          Mismatches:      210
Query Match:            8.85%           Indels:          123
DB:                     8               Gaps:            17

US-09-913-878A-2_COPY_710_1282 (1-573) x ATH011977 (1-1538)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
              |||:::|||::::::|||               |||:::::::::         |||||
Db        175 GGAATAGGAAAGATATCAGCTGAATTTGCTAGAAAGGTAGCGAGGAAATGTGGTCTTACA 234

Qy         21 AspVal---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMet-TrpValIl 39
              :::         |||||||||   |||   |||:::|||      ||||||:::  |||::::
Db        235 GAGTTTTCTCCATCTGCTTTTCAAATCCGTTATGGCGGTTATAAAGGAGTGGTGGCTGTT 294

Qy         39 eAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTr 59
              :
Db        295 G----------------------------------------------------------- 295

Qy         59 pGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGl 79
                          :::      |||||||||   |||   :::|||          |||::
Db        296 -------------ATCCAAACTCATCAAAGAAA-CTGTCTCTGAGGAAGAGTATGAGCAA 341

Qy         79 uLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeu------------------ProValLe 93
              :   :::|||          |||   :::|||                    |||
Db        342 ATTCGAATCGGAGAACACCAAGCTTGATGTTCTGGCGTGGAGCAAGTACCAACCTTGTTA 401

Qy         93 uGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAs 113
                 :::|||       :::                  :::   |||       :::
Db        402 TATGAACAGACAACTGATTACACTTTTGTCTACTCTTGGAGTTACGGACAGTGTGTTTGA 461

Qy        113 nAspLeuGlnArgGlnPheSerGluGlnLysHisAla---LeuAsnArgProValGluPh 132
                    ||||||:::       ::::::       |||   |||   |||:::|||
Db        462 GAAGAAACAAAGGGAAGTTGTGGATCGTCTCGACGCAATCTTGACTCATCCTTTGGAA-- 519

Qy        132 eArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgVa 152
                                                      :::|||         ::
Db        520 ---------------------------------------------GCTCACGAGGCTCT 533

Qy        152 lProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSe 172
              :      :::|||      |||             |||   |||      |||:::
Db        534 TGGTTTAATGGCT------CCAGGGGAAAACACAAATATTCTCAAGGCATTGATCTTGTG 587

Qy        172 rGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysAr 192
               |||:::   |||         :::   :::|||      :::      :::   :::
Db        588 TGGCTATAAACCTGACGCTGAACCTTTCCTTTCAATGATGCTTCAGAATTTCAGGGCATC 647

Qy        192 gLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTy 212
                |||         |||::::::|||         |||   :::             :::
Db        648 TAAGTTGTTGGAACTACGGACAAAAACAAGGATTTTCATTTCTGGTGGAAGATCTATGAT 707

Qy        212 rMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSe 232
                       |||                   ||||||   :::|||   |||   :::|||
Db        708 GGGATGCCTAGACGAGACCAGAACGCTGGAGTATGGTCAGGTAGTAGTGCAGTATTCAGA 767
```

```
Qy       232 rLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAl 252
                       |||                          :::           |||:::||||| 
Db       768 TCCCATGAGGCCGGGAAGGCGA---------TTCATCATCACCGGACCTGTTGTTGTTGC 818

Qy       252 aArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGl 272
             |::::::|||                  |||::::::  ::::::||||||      ||| 
Db       819 CAAAAACCCATGCCTGCATTCTGGTGACGTGCGTGTTCTTCAAGCTGTCAATGTCCCAGC 878

Qy       272 uLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLy 292
              |||:::   :::   |||   ::::::|||        ||||||      |||
Db       879 TTTAAATCATATGGTGGACTGCGTTGTGTTCCCGCAAAAGGGCTTGAGGCCACACCCAAA 938

Qy       292 sLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIl 312
              :::   ||||||  |||  |||||||||:::   :::||||||||||||  |||::
Db       939 TGAATGTTCTGGGAGTGATTTAGATGGAGATATTTACTTTGTATGTTGGGATCAAGAATT 998

Qy       312 eValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLy 332
              :|||              :::      |||       |||:::   ::::::   |||
Db       999 GGTTCCGCCAAGAACGTCTGAACCAATGGACTACACTCCTGAACCAACTCAAATCTTGGA 1058

Qy       332 sLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLy 352
                  |||
Db      1059 TCATGAT----------------------------------------------------- 1065

Qy       352 sGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPh 372
                         |||   :::   :::::::::   |||         :::      :::
Db      1066 -------GTCACAATTGAGGAAGTTGAAGAGTACTTTGCGAACTACATTGTGAATGATAG 1118

Qy       372 eLeuGlyMetCysThrAsnTyr------------------LysGluArgLeuCysTyrIl 386
              ||||||:::      |||                     ||||||   |||
Db      1119 TTTAGGGATCATCGCGAATGCTCATACCGCCTTTGCGGATAAGGAACCACTC-------- 1170

Qy       386 eAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVa 406
                          :::   |||   |||   |||:::                     ||
Db      1171 -------AAGGCGTTTAGTGACCCATGCATTGAGCTTGCAAAGAAGTTTTCAACTGCCGT 1223

Qy       406 lAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGl 426
              ||||      |||   |||:::                  |||   :::   :::
Db      1224 AGACTTTCCAAAAACTGGTGTTGCA------------------GCTGTGATACCGCAACA 1265

Qy       426 uLeu------------LeuGlyGlyAlaLeuSerLeuProAsp---ProMetTyrLysSe 441
              |||                            :::   ||||||   |||   |||:::||
Db      1266 TCTTTACGTGAAAGAGTATCCTGATTTCATGGAGAAACCGGATAAGCCAACATATGAGTC 1325

Qy       441 rAspSerTrpLeuGly-----------------------ArgGlyGluProThrHisIl 453
             |   :::   :::|||                        |||      ·|||
Db      1326 GAAGAACGTGATTGGTAAGCTTTTAGAGAGGTAAAAGAGCGAGCTCCACCGTTGATCTC 1385

Qy       453 eIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeu-------- 470
              |||                :::|||      :::   ||||||::::::
Db      1386 GATCAAATCATTCACGCTTGATGTGGCCTCAAAGTCTTATGATAAAGACATGGAAGTTGA 1445

Qy       471 ---------------------GluAlaPheHisAsnAlaMetLysAlaAlaLysAspTh 483
                                  |||||||||:::         ||||||      |||
Db      1446 CGGATTTGAGGAGTATGTTGATGAAGCTTTCTAC------CAGAAGGCGAATTATGATTT 1499

Qy       483 rGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyr 498
              :::   |||                  :::|||      ||||||
Db      1500 CAAGTTAGGT-----------------AATTTGATGGACTACTAT 1527
```

```
RESULT 7
AF468822
LOCUS       AF468822                6688 bp    DNA     linear   PLN 14-FEB-2002
DEFINITION  Diaporthe perjuncta putative RNA-dependent RNA polymerase RDP-1
            (rdp-1) gene, complete cds; and unknown genes.
ACCESSION   AF468822
VERSION     AF468822.1  GI:18656896
KEYWORDS    .
SOURCE      Diaporthe perjuncta
  ORGANISM  Diaporthe perjuncta
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Diaporthales; Valsaceae; Diaporthe.
REFERENCE   1  (bases 1 to 6688)
  AUTHORS   Preisig,O. and Wingfield,M.J.
  TITLE     Putative fungal RNA-dependent RNA polymerase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6688)
  AUTHORS   Preisig,O. and Wingfield,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2002) Forestry and Agricultural Biotechnology
            Institute, University of Pretoria, Lunnon Street 74, Pretoria 0002,
            South Africa
FEATURES             Location/Qualifiers
     source          1..6688
                     /organism="Diaporthe perjuncta"
                     /mol_type="genomic DNA"
                     /isolate="CMW3407"
                     /db_xref="taxon:186170"
                     /country="South Africa"
     mRNA            join(<1..275,348..>381)
                     /product="unknown"
     CDS             join(<1..275,348..381)
                     /note="similar to D-lactate dehydrogenase (D-lactate
                     ferricytochrome C oxidoreductase)"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAL78033.1"
                     /db_xref="GI:18656897"
                     /translation="MDDLGLFASILGHIGDGNFHESILYNPVNGEREKVEECVKKMVR
                     RAIDMDGTCTGEHSIGWGKKDSLLLEVGPETLAVMRAIKGSLDPRWIMNPGKIFDRQS
                     "
     gene            <1361..>5870
                     /gene="rdp-1"
     mRNA            join(<1361..4598,4654..>5870)
                     /gene="rdp-1"
                     /product="putative RNA-dependent RNA polymerase RDP-1"
     CDS             join(1361..4598,4654..5870)
                     /gene="rdp-1"
                     /codon_start=1
                     /product="putative RNA-dependent RNA polymerase RDP-1"
                     /protein_id="AAL78034.1"
                     /db_xref="GI:18656898"
                     /translation="MSSHHSTPTTPDRGPSAQPDASAQIHHTQTSSIPDATAKDNNPR
                     ISTLRDPAIADDTSDASNRNTQPSVVRRLPSPLSMSHTSNGNVLGQKVSLSGPIQGRS
                     IQTNASARTFGPRGHPNPRQNRFMPNRRQQRSPQSLQQRSTPAPLTGAVLSAAQSLTI
                     DLFGLPTDIKIREIWESMTSNHLNVNRVEIYENFDGKRTGKARVIIEPPPIRWPTWSG
                     PDNLITIDRADSSSLQVSLETRPFYPSTRTPCNRPIPAQFCQKKLLPTKLVFGMMASE
                     SAMLNMPQFSYEDVVLLTKFSKRQFELQFTVNLGRYGCKAFKSQIDFSHPTKIIWGRR
                     ENGAWALAMLLPSPPKLYGRLPVSHDDEQTLWQEWDLWARQTNISSEMHKLRDLPAKL
                     EKQTWKSIEFGRWTTYYLDMESSLSEWESVRTYLADYNVKILQVDHFDLVNPGLSNAW
                     KLLDNSAEMTLELLGSTQSYHLDFPVRYQLEVCISHGLLSEYGVNAQFLEKLISFETD
                     RARMMLEGVAEANKQFFEPMEMFDDPRILHYWPNARIPPYATLVRRAVITPTTIYFKT
                     PCVELTNRILRKYSDLNDRFLRVQFTDEVTFGKIFSSQDSNKDDNLYTRVHRVMQNGV
                     IIGDRHYKFLAFSNSQFRENGAFFFCETDHVTCASIRNWMGDFRHIRSVGKFAARMGQ
                     CFTTTRQVNGISIPKIRQIGDIQRRSGDNIWNFTDGVGKISVFFARMIASERDLPETP
                     SCFQMRMGGCKGVVVVWPDIPANEVHIRPSQEKFKAVYNGLEIIKTSTFAHATLNKQV
                     IPVLIALGVDDAVFVRMLDDELKEYDEALADSMKAGELLRSQVDENQTTLTMAEMVDT
                     FMDSEEPFLWTLLRLWKCWVLKRLKQKFAISVKNSAMIFGVVDETGVLRGHSQDTEGK
                     GFNNIESLPQIFLQVPIEGSDGKSTTNYEVITGICVVGRNPALHPGDVRVVEAVDVPE
                     LRHLKNVVVFPKTGDRDVPSMCSGGDMDGDDYFVYWDERLIPSEWDHPPLDHDADSST
                     PGLDSPADVTNGDVTRFFAQYMKNDSLGRIATAHFAQADQLAGGVKHPKCIELAKLHS
                     KAVDYIKSGKPAVMKRHLQPRNWPHWMERDKKSSYHSRGALGQIYDRIKIEEFHAAYE
                     MSFDTRILSRYQLEEDTLAKASEIKAAYDIAMRRLTGQHEAPVTEFEIWSTFILTKPR
                     VGSDYKLQENVGREMAALKARFRAVCMEAVTGDEQTPTCASASSKVNLEKLDRFVAAM
                     YTVTHNEVRAAVRERSKPRPDDEGDLEEIQMPLISFPWLYHRELARVALGRGGDVRPL
                     RKPVCADPAAANIESEPSGDGALSEEIDQEGETQLSLQKVRSKKDGYVIDRGESVETG
                     KVGEDYVRTSSGQVVRRGQLLTLFTDTGEEVQGAQSTIAPATPPAENNSGNPLVEDDG
                     TGSSYPASPRSEHEAEGEDEEIEFEEVDGANDGEEEDALEILARKIGM"
     mRNA            <6068..>6526
                     /product="unknown"
     CDS             6068..6526
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAL78035.1"
                     /db_xref="GI:18656899"
```

```
                    /translation="MVLQNHQQANRPTFLSLTYSSRASDYLISLVMAGDCFPLTWGPH
                    MWMQQPADQLGTSPGTPSRDDPFWHGQRDSQHRAFFQLAPGTSLDVLDWIGLESAEIP
                    ARLYSSLSSIHIRLSPSRPSTMDSRAVSSSLRTADDSQKVAHCQLFMLPF"
ORIGIN

Alignment Scores:
Pred. No.:               2.84e-12       Length:         6688
Score:                   265.50         Matches:        137
Percent Similarity:      38.27%         Conservative:   67
Best Local Similarity:   25.70%         Mismatches:     199
Query Match:             8.82%          Indels:         131
DB:                      8              Gaps:           23

US-09-913-878A-2_COPY_710_1282 (1-573) x AF468822 (1-6688)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
                |||||||:::::|||             |||:::   |||                |||
Db       3512 GGAGTGGGCAAGATATCCGTCTTCTTTGCAAGAATGATTGCAAGCGAGCGTGATTTGCCA 3571

Qy         21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
              :::   |||||      |||   |||   |||         ||||||:::   |||:::
Db       3572 GAGACGCCTTCATGCTTTCAGATGAGAATGGGCGGGTGCAAGGGCGTGGTCGTTGTATGG 3631

Qy         41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
                 |||         :::|||      :::         |||||||||   |||::::::
Db       3632 CCTGACATCCCAGCGAACGAG------GTCCATATACGACCTTCTCAGGAGAAGTTTAAG 3685

Qy         61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
                    :::                           ||||||:::         |||
Db       3686 GCCGTGTAC---------------AATGGTCTCGAGATCATCAAG---ACCTCTACGTTC 3727

Qy         81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
                    |||   ||||||   |||::::::|||||||||                  |||
Db       3728 GCCCATGCTACTCTCAATAAGCAGGTTATACCTGTCCTGATCGCCCTGGGTGTGGAT--- 3784

Qy        101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
                                |||:::      |||:::::::::|||      :::::::::
Db       3785 ------------GACGCCGTTTTCGTCAGAATGCTCGATGATGAGCTCAAAGAGTACGAC 3832

Qy        121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
               |||         ||||||                                    :::
Db       3833 GAA---------GCCTTA-------------------------------------GCTGAT 3847

Qy        141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
              |||   :::|||      :::   |||                            |||
Db       3848 TCAATGAAAGCCGGTGAACTTCTTCGGTCTCAAGTC---------------------GAC 3886

Qy        161 SerGlnGluGluThrLeuAsnPheLeu---MetAsnSerGlyPheAspProLysLysGln 179
                    :::   ||||||         |||         |||         ::::::
Db       3887 GAGAACCAAACAACATTGACCATGGCGGAGATGGTCGACACGTTCATGGACTCCGAAGAG 3946

Qy        180 LysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArg---LysCys---------Asp 195
                 :::|||               |||   |||   :::      ||||||
Db       3947 CCGTTCCTC-----------TGGACCTTGCTTCGTTTGTGGAAGTGCTGGGTGTTGAAG 3994

Qy        196 ThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAla 215
                 ||||||   |||         |||   |||      ||||||   |||:::   :::
Db       3995 AGGTTAAAGCAAAAGTTCGCCATCAGCGTCAAGAACAGCGCTATGATCTTCGGGGTTGTC 4054

Qy        216 AspPheTrpGlyValLeuGluGluAsn--------------------------------- 224
              |||      |||||||||      :::
Db       4055 GATGAGACCGGTGTGCTGAGGGGCCACTCACAAGACACCGAAGGCAAAGGGTTCAATAAC 4114

Qy        225 ---------------------------GluValHisVal---GlyPheSerSerLysPhe 234
                                         :::|||   :::   |||         |||
Db       4115 ATCGAGTCGCTCCCACAGATCTTTCTCCAAGTCCCTATCGAGGGCAGTGATGGGAAGAGC 4174

Qy        235 ArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSer 254
                          |||      |||   :::                  :::|||   |||:::
Db       4175 ACCACGAATTACGAGGTCATCACTGGAATT-----------TGCGTAGTAGGCCGTAAC 4222

Qy        255 ProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHis 274
              ||||||         |||   |||::::::   |||   ||||||      |||||||||

Db       4223 CCAGCGCTTCATCCTGGAGATGTGCGGGTTGTTGAGGCAGTTGACGTGCCAGAGTTACGA 4282

Qy        275 SerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeu 294
                 ||||||:::|||::::::|||            ||||||         :::
Db       4283 CATCTCAAGAACGTGGTCGTCTTTCCAAAGACTGGAGACAGGGACGTTCCGAGCATGTGC 4342

Qy        295 SerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleVal--- 313
              ||||||||||||   |||||||||         :::|||   ||||||      ::::::
Db       4343 TCCGGTGGTGACATGGACGGAGATGACTACTTTGTTTACTGGGACGAGCGACTCATCCCA 4402

Qy        314 -----------------------AspGlyPheValAsnAlaGluMetPro--------- 322
                                     |||                  |||
Db       4403 AGCGAGTGGGACCACCCACCTCTGGACCACGATGCGGATTCATCGACTCCTGGGCTGGAC 4462

Qy        323 ---------------LeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThr 337
                                 |||::::::|||:::
Db       4463 AGCCCGGCAGATGTGACAAACGGGGACGTCACTAGGTTC-------------------- 4501

Qy        338 PheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyr 357
              |||   |||   |||   :::         |||      |||
Db       4502 TTTGCCCAGTATATGAAGAACGATTCCCTTGGCCGTATCGCC------------------ 4543

Qy        358 AspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGly--------- 374
                          :::   |||||||||                  |||
Db       4544 ---------------ACAGCGCACTTTGCTCAGGCAGATCAGTTGGCAGGGGGCGTCAAG 4588

Qy        375 ------MetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsn 392
                    |||               ||||||:::          ::::::
Db       4589 CACCCAAAATGTAAGTAAATCGCTACGACCTTGTGCTGGC--CGGGGTGAACTGACGGCT 4646

Qy        393 LysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGly 412
              :::      |||   |||:::   |||               ||||||   |||   |||
Db       4647 GAAGTAGGCATCGAGCTCGCCAAGCTCCACTCAAAAGCTGTCGACTATATCAAGTCCGGC 4706

Qy        413 IleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeu 432
                                |||   ::::::|||   |||
Db       4707 ------------------AAACCCGCGGTCATGAAGAGACATCTT------CAGCCTCGA 4742

Qy        433 SerLeuProAspProMetTyr------LysSerAspSerTrpLeuGlyArgGlyGluPro 450
              :::   |||      |||         |||      |||::::    ||||||
Db       4743 AACTGGCCGCATTGGATGGAGAGGGACAAGAAGTCCTCCTACCACTCCAGAGGCGCGCTG 4802

Qy        451 ThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeu 470
                    |||   |||   :::|||                                 :::
Db       4803 GGTCAAATTTACGACCGGATCAAG--------------------------------ATC 4829

Qy        471 GluAlaPheHisAsnAlaMetLysAlaAlaLysAspThr 483
              |||   ||||||   |||   :::   :::   ||||||
Db       4830 GAGGAATTCCATGCCGCTTACGAAATGTCTTTCGACACG 4868

RESULT 8
DDI314910
LOCUS       DDI314910               5992 bp    DNA     linear   INV 14-JUN-2001
DEFINITION  Dictyostelium discoideum partial RrpB gene for putative RNA
            dependent RNA polymerase, exons 1-2.
ACCESSION   AJ314910
VERSION     AJ314910.1  GI:14475570
KEYWORDS    RNA dependent RNA polymerase; RrpB gene.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1
  AUTHORS   Martens,H., Novotny,J., Oberstrass,J., Steck,T.L., Postlethwait,P.
            and Nellen,W.
  TITLE     RNAi in Dictyostelium: developmental regulation and the role of
            RdRPs and dsRNase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5992)
  AUTHORS   Martens,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2001) Martens H., Genetics Department, Kassel
            University, Heinrich-Plett Strasse 40, 34132, GERMANY
FEATURES             Location/Qualifiers
```

```
     source         1. .5992
                    /organism="Dictyostelium discoideum"
                    /mol_type="genomic DNA"
                    /strain="AX4"
                    /db_xref="taxon:44689"
     gene           1. .5992
                    /gene="RrpB"
     mRNA           join(<1. .970,1067. .5992)
                    /gene="RrpB"
     CDS            join(<1. .970,1067. .5992)
                    /gene="RrpB"
                    /codon_start=2
                    /product="putative RNA dependent RNA polymerase"
                    /protein_id="CAC41975.1"
                    /db_xref="GI:14475571"
                    /db_xref="GOA:Q95ZG6"
                    /db_xref="SPTREMBL:Q95ZG6"
                    /translation="RLTIEIVQRSIKDMEERMLSRVFKPHSLLTNSTQSELQPELVSF
                    KTSGQESLIETSICEFLVSNSKELQIFLNYSDVDLDGVRGNPSFLKALERLDKYSKQL
                    ESIYIEYTKVLIKLYELLSVLSTEGPKQVLTGLDLLMQSTNQENELYNKLLELRYYIS
                    GQYEDRESNFEKGSTRYRKLISSLEYAISDASDGELQKDLRILVFVETRSGASNLTSM
                    LKKEPFQEYLHTKRLVGHNGDDGMDSEKQQSIIRKFRDGKCRLIVTTNVLEEGIDVQD
                    CNIVICYDGILSLKSLIQRRGRARSKNESKFIIIYNDDKDLKILNILSSENLLNNSIN
                    QILSNRERQAITKNQFLNWKKLNINNEYFINNQLDGGPLDKDIKCVILSIYHIEIDIE
                    TFRELLLFQLPDIVNQVDQVSIRMEGDDEYLSSSIIVILNTCIDKEQSIYRLKSLLNI
                    NGLKYWMIDRSSFHFEDDNIDNDNNNSYDRENEQDNFNYFINEYTHDLSQFKNKNYSI
                    GSFEYGNMSSPTKYHGIGQLDDSFDLFYGNREMFFTKKTRFQRYRLNLKLTDIESTFL
                    LVNDKYQNIVHILLIVKRPLIFEEEKETTHRVFWNKSKPKKDDFGQAFVYRLTLSTID
                    NEIKEVIKPLIANGFRMYESSLELNCFIQGGINKNNNCMIIDLNLQYIKRNLEIHYLL
                    QCLQSQSSYGRVLNQDFIVEINNLLAFGKQEEAEFIITSLISKRTKFINYKEFIGDCL
                    KNFLPDPYYAKKKDLSTKSGEFTMIRDVFITPSRTIFIQPTLQRSCRAVRKFGSSNFI
                    MVKIVNESLEPLQGVNNKHPLHENRIKPILEGGILVGGKIYSYAGNSNSQLREYSSWF
                    VSNQIGTHTVKIWSGIEHVDNVRKFFRCIGLMFSTTIPTVTLPQNRIYRIQDITRNTH
                    VFTEGCGEIGPELAKHLNENYNFRPSTCAYQVRIGGNKGMLVVNNQAPDPSGIYIRPS
                    MVKFNPIVCGDEHRTLEICSVSTTSRCKLNRQVISLLSTLGTQDNVFFALQDHYLNQV
                    AQIVNDTNASKQAIVEFFPDITEGELYQDPYIRRILISLYKLKMERIQQKCHIEIKDS
                    RMLLGVCDPTNSLPPNTVFVQLEEEDEDDDDGRKYEKVIEGLVMVIKNPCTHPGDVRY
                    LKAVDNLRLRHLRNVLVFSTKGDVPNFKEISGSDLDGDRYFFCYDKSLIGNRSKSETA
                    YLGDETVSNNDKKANVFNDPFALSSMYSTNVERQELGKMYNSHLAISDLFGANHEFSI
                    QISKECFKEIDYPKTGIHGTIPKEANVWLKTKGYPHYMQRENSTRVYYQSKTIMGKMY
                    DQIDQLVYIGDFLPNISLDKSNLVDDYEIYLNSAKILYSQYKLQVHSLLRHYSAESEE
                    SIMIGFLDQGFISDKVSKDIKGEMKNDYIKIQQTFENEFLKEFGEQHKENCLLIHRVN
                    IEKKVSAWYHVAYSDLKDDRALGFYWIARSFKQMTLDQESQKKKNILVNSILDHMNKK
                    KDTLFPLYESRLMIVKELTLFLNMLNDTVFKESTLEIIGSTATMLFDENSNLNLYLKL
                    ESQNDHMNLAEKNEILNQLKDEIKTSSFISAEACSIIKIQGDIPLFKFKIDKIQCLIG
                    LETSKWHETLVQQQYITKNPSILSILFSIIDWSINSLFVDSETGEIINNSKVYSSLNL
                    QKVIVENSMIPSIFNNKNHNNNNNNNNNNKHSVKNLTKSNLFSLLIDFCIINKYIDKC
                    NPNELTNTTSNLTIGNWVKLIEIASRYFSVDNNNNNNNKGHKRKKNKNLDEYNTNNLG
                    DTILSFFRFYSDSFGRSIYGNEPFIIKLFSKDANDIENQILNPSTESHKTALLFYQES
                    FRHAFHSLSTFVSVEEFLSNCLKTTKKIMPIKKNQINLLQERSVNEEIVKNVQLCGCS
                    IKIDTNVKNPYILIDGKPQQIRLAQNAIRDLLSIKTENYNLAITKRNNKRKNFTLGDF
                    AKNKIK"
     exon           <1. .970
                    /gene="RrpB"
                    /number=1
     intron         971. .1066
                    /gene="RrpB"
                    /number=1
     exon           1067. .5992
                    /gene="RrpB"
                    /number=2
ORIGIN

Alignment Scores:
Pred. No.:              4e-12          Length:          5992
Score:                  263.00         Matches:         134
Percent Similarity:     36.53%         Conservative:    91
Best Local Similarity:  21.75%         Mismatches:      209
Query Match:            8.74%          Indels:          182
DB:                     3              Gaps:            25

US-09-913-878A-2_COPY_710_1282 (1-573) x DDI314910 (1-5992)
```

```
Qy       1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
           |||   |||   :::         :::||||||   :::   :::
Db    2852 GGATGTGGTGAAATTGGTCCAGAACTTGCTAAACATTTAAATGAAAATTATAATTTCAGA 2911

Qy      21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
                       |||   |||   |||   |||      |||||||||   |||::::::
Db    2912 CCATCAACATGTGCATATCAAGTTAGAATAGGTGGAAATAAAGGTATGTTGGTGGTAAAC 2971

Qy      41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
              :::         |||         |||         ||||||      |||:::
Db    2972 ---AATCAAGCACCCGATCCATCTGGTATTTATATTAGACCAAGTATGGTGAAATTCAAT 3028

Qy      61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
                                   ||||||||||||:::   ||||||   |||
Db    3029 CCAATCGTTTGTGGTGATGAACATAGAACATTAGAGATTTGCAGTGTT---TCAACAACT 3085

Qy      81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAsp---ArgAlaArgAsp 99
                       ||||||   |||::::::   :::|||               :::|||
Db    3086 TCAAGATGTAAATTAAATAGACAAGTAATTAGTTTACTCTCAACATTGGGTACACAAGAT 3145

Qy     100 LysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPhe 119
              |||                                                   |||
Db    3146 AATGTT------------------------------------------------TTCTTT 3157

Qy     120 SerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSer 139
           :::   |||   |||   ||||||         :::      |||   |||   ::::::
Db    3158 GCATTACAAGATCATTATCTAAAT---------CAAGTGGCTCAAATCGTAAATGATACC 3208

Qy     140 TyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuPro 159
              :::|||:::         |||                                 |||
Db    3209 AATGCTTCCAAACAAGCAATCGTTGAATTC------------------------TTTCCA 3244

Qy     160 AspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGln 179
           |||      |||      |||                                 :::
Db    3245 GATATTACAGAGGGTGAACTT-----------------------------TATCAAGAT 3274

Qy     180 LysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSer 199
              |||::::::   |||         |||   |||   |||   :::   ::::::
Db    3275 CCTTACATTCGTAGAATTTTAATTAGTCTTTATAAATTAAAAATGGAGAGAATTCAACAA 3334

Qy     200 LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219
           |||   :::|||   :::      |||      :::      :::   |||
Db    3335 AAATGTCATATAGAGATTAAAGATAGTAGAATGTTGTTGGGTGTTTGTGATCCAACAAAT 3394

Qy     220 ValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGlu 239
              |||      |||   |||   |||            :::      |||:::::::::
Db    3395 TCATTACCACCAAATACTGTATTTGTTCAATTGGAAGAGGAAGATGAAGATGATGATGAT 3454

Qy     240 ------SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHis 257
                    :::      :::   :::      |||:::|||   ::::::|||
Db    3455 GGTCGTAAATATGAAAAAGTAATTGAAGGATTGGTAATGGTAATTAAAAATCCTTGTACT 3514

Qy     258 PheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLys 277
              |||   |||::::::   ::::::||||||            |||      |||:::
Db    3515 CATCCTGGTGATGTAAGATATTTAAAGGCTGTTGATAATTTAAGATTAAGACATTTAAGA 3574

Qy     278 AspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGly 297
           :::|||::::::||||||||||||||||||||||||      |||::::::||||||
Db    3575 AATGTATTAGTTTTCTCAACAAAAGGTGACGTACCAAACTTTAAAGAAATTTCAGGTTCA 3634

Qy     298 AspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheVal 317
           |||   |||||||||      :::   |||:::|||      ::::::
Db    3635 GATTTAGATGGTGATCGTTATTTCTTTTGTTACGATAAATCATTGATT------------ 3682

Qy     318 AsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThr 337

Db    3682 ------------------------------------------------------------ 3682

Qy     338 PheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyr 357
                                         |||:::   :::|||   :::|||   |||
Db    3683 ------------------------------GGTAATCGTTCAAAATCAGAGACAGCCTAT 3712

Qy     358 AspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGly--------- 374
```

```
Db       3713 ---------------------------------------------TTAGGTGACGAGACC 3727

Qy        375 MetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysPro 394

Db       3728 GTTTCAAACAATGATAAAAAGGCC-----------------AATGTTTTCAATGATCCA 3769

Qy        395 AlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleVal 414

Db       3770 ---TTTGCTTTATCCTCAATGTACTCTACAAATGTTGAACGTCAAGAACTTGGTAAAATG 3826

Qy        415 PheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAla--------- 431

Db       3827 TATAAT------------TCACATTTGGCAATCTCTGATCTTTTCGGTGCAAATCACGAA 3874

Qy        432 ------LeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArgGlyGlu 449

Db       3875 TTTTCAATTCAAATCTCAAAAGAATGTTTTAAAGAA------------------------ 3910

Qy        450 ProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGlu 469

Db       3911 ------------ATTGATTATCCAAAAACTGGTATT---CATGGCACTATTCCAAAAGAA 3955

Qy        470 LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPhe 489

Db       3956 GCAAATGTTTGG--------------------TTAAAAACTAAAGGTTATCCACATTAT 3994

Qy        490 TrpAspProAspLeuAlaSer------TyrTyr------------------------- 498

Db       3995 ATGCAAAGAGAGAATTCAACCAGAGTTTATTATCAATCGAAAACCATTATGGGTAAAATG 4054

Qy        499 --------------------------------ThrPhePheLysGluIleSer---Asp 506

Db       4055 TATGATCAAATTGATCAGTTGGTTTATATTGGTGATTTCTTACCAAATATTAGTCTTGAT 4114

Qy        507 LysSer------------------------------ArgSerSerAlaLeuLeuPheThr 516

Db       4115 AAAAGCAATTTGGTTGATGACTATGAAATCTATTTAAATAGTGCTAAAATCTTATATAGT 4174

Qy        517 ThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArg--------------- 531

Db       4175 CAGTATAAACTTCAAGTTCATTCATTGTTAAGACATTATTCAGCGGAGTCTGAAGAATCT 4234

Qy        532 --------------------------LeuValLysAsnLysGluMetArgAsp------ 540

Db       4235 ATAATGATTGGTTTTCTTGATCAAGGTTTTATTAGTGACAAAGTTTCAAAAGATATCAAA 4294

Qy        541 ------SerLysAspProTyrProValArgValAsnGlnValTyrGlu 554

Db       4295 GGTGAAATGAAAAATGATTAC---ATTAAAATTCAACAAACTTTTGAA 4339
```

RESULT 9
AR145904

```
LOCUS       AR145904             3731 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 1 from patent US 6218142.
ACCESSION   AR145904
VERSION     AR145904.1  GI:15109093
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3731)
  AUTHORS   Wassenegger,M., Riedel,L., Schiebel,W. and Sanger,H.L.
  TITLE     Nucleic acid molecules encoding polypeptides having the enzymatic
            activity of an RNA-directed RNA polymerase (RDRP)
  JOURNAL   Patent: US 6218142-A 1 17-APR-2001;
FEATURES             Location/Qualifiers
     source          1..3731
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

Alignment Scores:

```
Pred. No.:              2.6e-12     Length:         3731
Score:                  262.00      Matches:        127
Percent Similarity:     37.66%      Conservative:   85
Best Local Similarity:  22.56%      Mismatches:     222
Query Match:            8.70%       Indels:         130
DB:                     6           Gaps:           17
```

US-09-913-878A-2_COPY_710_1282 (1-573) x AR145904 (1-3731)

```
Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20

Db       1772 GGAATTGGTAAAATATCTGGTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCTTCAA 1831

Qy         21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40

Db       1832 TATACCCCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGGTGTTGAT 1891

Qy         41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60

Db       1892 CCGGATTCATCAATGAAG---------TTGTCTTTGAGAAAGAGCATGTCGAAATATGAA 1942

Qy         61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80

Db       1943 TCAGACAACATA---------------AAGTTAGATGTCCTTGGA---TGGAGCAAATAT 1984

Qy         81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100

Db       1985 CAGCCTTGTTATCTTAATCGTCAACTGATTACGCTCTTGTCT------------------ 2026

Qy        101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120

Db       2027 ---------ACACTTGGAGTGAAAGATGAAGTTCTCGAACAGAAGCAAAAGGAAGCTGTA 2077

Qy        121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140

Db       2078 GATCAGCTTGATGCTATCTTGCATGATTCTTTGAAGGCACAGGAGGCTTTGGAATTGATG 2137

Qy        141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160

Db       2138 TCT---------------------------------------------------CCTGGA 2146

Qy        161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180

Db       2147 GAGAACACTAATATTCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCC 2206

Qy        181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200

Db       2207 TTTCTTTCAATGATGTTGCAAACCTTCCGCGCATCCAAGTTGCTCGATTTGCGGACTAGA 2266

Qy        201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220

Db       2267 TCAAGAATATTTATTCCAAATGGAAGAACAATGATGGGATGTTTGGATGAATCCAGAACC 2326

Qy        221 LeuGluGluAsnGluValHisValGlyPheSer------------SerLysPheArgAsp 236

Db       2327 TTGGAATATGGTCAGGTGTTTGTTCAGTTTACTGGTGCTGGACATGGAGAGTTTTCTGAC 2386

Qy        237 GluGluGluSerPhe-----------------------ThrLeuLeuSerAspCysAsp 248

Db       2387 GATTTACATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAAT 2446

Qy        249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268

Db       2447 GTGGTTGTTGCAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTAAAGGCTGTA 2506

Qy        269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288

Db       2507 AATGTTCGAGCGCTGCACCACATGGTAGATTGTGTTGTATTCCCTCAGAAAGGAAAAAGA 2566

Qy        289 ProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrp 308

Db       2567 CCTCATCCGAATGAATGTTCTGGGAGTGATTTGGATGGGGATATCTACTTTGTTTGCTGG 2626
```

```
Qy        309 AspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSer 328
              |||    ::::::::                    |||       |||     |||
Db       2627 GATCAAGACATGATCCCGCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCCAGC 2686

Qy        329 ArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGly 348
                    |||       |||
Db       2687 ATACAGTTGGACCATGAT------------------------------------------ 2704

Qy        349 SerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeu 368
                                   |||   :::   ::::::::   |||        :::
Db       2705 -----------------GTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATT 2746

Qy        369 GlnProAsnPheLeuGlyMetCysThrAsn------------TyrLysGluArgLeuCys 384
                :::    |||||||:::      |||            :::    :::|||
Db       2747 GTGAATGACAGTTTGGGAATCATAGCAAATGCCCATGTCGTATTTGCAGACAGA------ 2800

Qy        385 TyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
                        :::        :::  |||         |||:::   |||
Db       2801 ---GAACCTGATATGGCCATGAGTGATCCATGCAAAAAACTTGCTGAGCTCTTTTCAATT 2857

Qy        405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
                 ||||||      |||   |||:::                |||      |||||||||
Db       2858 GCAGTGGACTTTCCAAAGACTGGTGTTCCCGCTGAAATACCATCT------CAGTTGCGC 2911

Qy        425 ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTr 444
                                                     |||      |||  |||||
Db       2912 ------------------------------------CCTAAAGAATACCCAGACTTCATG 2935

Qy        444 pLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgPr 464
              |:::   |||                                          |||||
Db       2936 GATAAGCCGG---------------------------------------ACAAGACC 2953

Qy        464 oAlaIleAspLysGlu------LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAs 482
               ||||||   ::::::       ||||||:::|||          :::|||
Db       2954 AGCTATATCTCAGAAAGAGTTATTGGAAAGCTTTTC-AGGAAAGTGAAG---------- 3001

Qy        482 pThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLy 502
                    |||   |||           |||       :::|||    ||||||    ::
Db       3002 -------GACAAAGCA-----------CCTCAGGCTAGCTCTATCGCGACCTTCACAAG 3042

Qy        502 sGluIleSerAspLysSerArgSerSerAlaLeu------------------------ 513
              :::::::::   :::|||    :::   :::
Db       3043 AGATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACAT 3102

Qy        514 ----------------LeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGl 528
                                  |||        |||::::::|||    :::
Db       3103 TGACGAAGCTTTTGACTACAAAACTGAATATGACAACAAGCTGGGTAATTTAATGGACTA 3162

Qy        528 uTyrGly 530
               ||||||
Db       3163 CTATGGC 3169
```

```
RESULT 10
LERDRP
LOCUS       LERDRP                 3731 bp    mRNA    linear   PLN 18-DEC-1998
DEFINITION  L.esculentum mRNA for RNA-directed RNA polymerase.
ACCESSION   Y10403
VERSION     Y10403.1  GI:4038591
KEYWORDS    RdRP gene; RNA-directed RNA polymerase.
SOURCE      Lycopersicon esculentum (tomato)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1
  AUTHORS   Schiebel,W., Pelissier,T., Riedel,L., Thalmeir,S., Schiebel,R.,
            Kempe,D., Lottspeich,F., Sanger,H.L. and Wassenegger,M.
  TITLE     Isolation of an RNA-directed RNA polymerase-specific cDNA clone
            from tomato
  JOURNAL   Plant Cell 10 (12), 2087-2101 (1998)
  MEDLINE   99055198
   PUBMED   9836747
REFERENCE   2  (bases 1 to 3731)
  AUTHORS   Wassenegger,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-1997) M. Wassenegger, Max-Planck-inst. fuer
            Biochemie, Viroidforschung, Am Klopferspitz 18a,
            Planegg-Martinsried, 82152, FRG
FEATURES             Location/Qualifiers
     source          1..3731
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /strain="Rutgers"
                     /db_xref="taxon:4081"
     gene            1..3731
                     /gene="RdRP"
     CDS             194..3538
                     /gene="RdRP"
                     /EC_number="2.7.7.48"
                     /codon_start=1
                     /product="RNA-directed RNA polymerase"
                     /protein_id="CAA71421.1"
                     /db_xref="GI:4038592"
                     /db_xref="GOA:Q9ZR58"
                     /db_xref="SPTREMBL:Q9ZR58"
                     /translation="MGKTIQVFGFPYLLSAEVVKSFLEKYTGYGTVCALEVKQSKGGS
                     RAFAKVQFADNISADKIITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGITLNF
                     GCQISDDKFAVLGSTEVSIQFGIGLKKFFFFLSSGSADYKLQLSYENIWQVVLHRPYG
                     QNAQFLLIQLFGAPRIYKRLENSCYSFFKETPDDQWVRTTDFPPSWIGLSSSLCLQFR
                     RGVRLPNFEESFFHYAERENNITLQTGFTFFVSQKSALVPNVQPPEGISIPYKILFKI
                     SSLVQHGCIPGPALNVYFFRLVDPRRRNVACIEHALEKLYYIKECCYDPVRWLTEQYD
                     GYLKGRQPPKSPSITLDDGLVYVRRVLVTPCKVYFCGPEVNVSNRVLRNYSEDIDNFL
                     RVSFVDEEWEKLYSTDLLPKASTGSGVRTNIYERILSTLRKGFVIGDKKFEFLAFSSS
                     QLRDNSVWMFASRPGLTANDIRAWMGDFSQIKNVAKYAARLGQSFGSSRETLSVLRHE
                     IEVIPDVKVHGTSYVFSDGIGKISGDFAHRVASKCGLQYTPSAFQIRYGGYKGVVGVD
                     PDSSMKLSLRKSMSKYESDNIKLDVLGWSKYQPCYLNRQLITLLSTLGVKDEVLEQKQ
                     KEAVDQLDAILHDSLKAQEALELMSPGENTNILKAMLNCGYKPDAEPFLSMMLQTFRA
                     SKLLDLRTRSRIFIPNGRTMMGCLDESRTLEYGQVFVQFTGAGHGEFSDDLHPFNNSR
                     STNSNFILKGNVVVAKNPCLHPGDIRVLKAVNVRALHHMVDCVVFPQKGKRPHPNECS
                     GSDLDGDIYFVCWDQDMIPPRQVQPMEYPPAPSIQLDHDVTIEEVEEYFTNYIVNDSL
                     GIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDFPKTGVPAEIPSQLRPKEYPDFM
                     DKPDKTSYISERVIGKLFRKVKDKAPQASSIATFTRDVARRSYDADMEVDGFEDYIDE
                     AFDYKTEYDNKLGNLMDYYGIKTEAEILSGGIMKASKTFDRRKDAEAISVAVRALRKE
                     ARAWFKRRNDIDDMLPKASAWYHVTYHPTYWGCYNQGLKRAHFISFPWCVYDQLIQIK
                     KDKARNRPVLNLSSLRAQLSHRLVLK"
ORIGIN

Alignment Scores:
Pred. No.:              2.6e-12      Length:          3731
Score:                  262.00       Matches:         127
Percent Similarity:     37.66%       Conservative:    85
Best Local Similarity:  22.56%       Mismatches:      222
Query Match:            8.70%        Indels:          130
DB:                     8            Gaps:            17

US-09-913-878A-2_COPY_710_1282 (1-573) x LERDRP (1-3731)
```

```
Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
              |||:::|||:::::::|||             |||   |||:::             ||||||
Db       1772 GGAATTGGTAAAATATCTGGTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCTTCAA 1831

Qy         21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
                 |||||||||   |||   |||:::|||        ||||||:::        :::|||
Db       1832 TATACCCCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGGTGTTGAT 1891

Qy         41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
                 |||   :::                 :::         |||      |||:::|||
Db       1892 CCGGATTCATCAATGAAG---------TTGTCTTTGAGAAAGAGCATGTCGAAATATGAA 1942

Qy         61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
                 |||   :::               |||:::|||            |||:::
Db       1943 TCAGACAACATA---------------AAGTTAGATGTCCTTGGA---TGGAGCAAATAT 1984

Qy         81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
              :::         ||||||   ||||||:::   :::|||
Db       1985 CAGCCTTGTTATCTTAATCGTCAACTGATTACGCTCTTGTCT------------------ 2026
```

```
Qy        101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
Db       2027 ---------ACACTTGGAGTGAAAGATGAAGTTCTCGAACAGAAGCAAAAGGAAGCTGTA 2077

Qy        121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
Db       2078 GATCAGCTTGATGCTATCTTGCATGATTCTTTGAAGGCACAGGAGGCTTTGGAATTGATG 2137

Qy        141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
Db       2138 TCT-------------------------------------------------CCTGGA 2146

Qy        161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
Db       2147 GAGAACACTAATATTCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCC 2206

Qy        181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
Db       2207 TTTCTTTCAATGATGTTGCAAACCTTCCGCGCATCCAAGTTGCTCGATTTGCGGACTAGA 2266

Qy        201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
Db       2267 TCAAGAATATTTATTCCAAATGGAAGAACAATGATGGGATGTTTGGATGAATCCAGAACC 2326

Qy        221 LeuGluGluAsnGluValHisValGlyPheSer-----------SerLysPheArgAsp 236
Db       2327 TTGGAATATGGTCAGGTGTTTGTTCAGTTTACTGGTGCTGGACATGGAGAGTTTTCTGAC 2386

Qy        237 GluGluGluSerPhe-----------------------ThrLeuLeuSerAspCysAsp 248
Db       2387 GATTTACATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAAT 2446

Qy        249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
Db       2447 GTGGTTGTTGCAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTTAAAGGCTGTA 2506

Qy        269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288
Db       2507 AATGTTCGAGCGCTGCACCACATGGTAGATTGTGTTGTATTCCCTCAGAAAGGAAAAAGA 2566

Qy        289 ProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrp 308
Db       2567 CCTCATCCGAATGAATGTTCTGGGAGTGATTTGGATGGGGATATCTACTTTGTTTGCTGG 2626

Qy        309 AspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSer 328
Db       2627 GATCAAGACATGATCCCGCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCCAGC 2686

Qy        329 ArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGly 348
Db       2687 ATACAGTTGGACCATGAT------------------------------------------ 2704

Qy        349 SerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeu 368
Db       2705 ------------------GTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATT 2746

Qy        369 GlnProAsnPheLeuGlyMetCysThrAsn-----------TyrLysGluArgLeuCys 384
Db       2747 GTGAATGACAGTTTGGGAATCATAGCAAATGCCCATGTCGTATTTGCAGACAGA------ 2800

Qy        385 TyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
Db       2801 ---GAACCTGATATGGCCATGAGTGATCCATGCAAAAAACTTGCTGAGCTCTTTTCAATT 2857

Qy        405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
Db       2858 GCAGTGGACTTTCCAAAGACTGGTGTTCCCGCTGAAATACCATCT------CAGTTGCGC 2911

Qy        425 ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTr 444
Db       2912 -------------------------------------CCTAAAGAATACCCAGACTTCATG 2935

Qy        444 pLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgPr 464
Db       2936 GATAAGCCGG-------------------------------------------ACAAGACC 2953

Qy        464 oAlaIleAspLysGlu------LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAs 482
Db       2954 AGCTATATCTCAGAAAGAGTTATTGGAAAGCTTTTC-AGGAAAGTGAAG----------- 3001

Qy        482 pThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLy 502
Db       3002 -------GACAAAGCA------------CCTCAGGCTAGCTCTATCGCGACCTTCACAAG 3042

Qy        502 sGluIleSerAspLysSerArgSerSerAlaLeu------------------------- 513
Db       3043 AGATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACAT 3102

Qy        514 ----------------LeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGl 528
Db       3103 TGACGAAGCTTTTGACTACAAAACTGAATATGACAACAAGCTGGGTAATTTAATGGACTA 3162

Qy        528 uTyrGly 530
Db       3163 CTATGGC 3169

RESULT 11
DDI314909
LOCUS       DDI314909              5731 bp    DNA     linear   INV 14-JUN-2001
DEFINITION  Dictyostelium discoideum partial RrpA gene for putative RNA
            dependent RNA polymerase, exons 1-2.
ACCESSION   AJ314909
VERSION     AJ314909.1  GI:14475568
KEYWORDS    RNA dependent RNA polymerase; RrpA gene.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1
  AUTHORS   Martens,H., Novotny,J., Oberstrass,J., Steck,T.L., Postlethwait,P.
            and Nellen,W.
  TITLE     RNAi in Dictyostelium: developmental regulation and the role of
            RdRPs and dsRNase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5731)
  AUTHORS   Martens,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2001) Martens H., Genetics Department, Kassel
            University, Heinrich-Plett Strasse 40, 34132, GERMANY
FEATURES             Location/Qualifiers
     source          1. .5731
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
     gene            1. .5731
                     /gene="RrpA"
     mRNA            join(<1. .1572,1655. .>5731)
                     /gene="RrpA"
     CDS             join(1. .1572,1655. .>5731)
                     /gene="RrpA"
                     /function="required for RNA interference"
                     /codon_start=1
                     /product="putative RNA dependent RNA polymerase"
                     /protein_id="CAC41974.1"
                     /db_xref="GI:14475569"
                     /db_xref="GOA:Q95ZG7"
                     /db_xref="SPTREMBL:Q95ZG7"
                     /translation="METKDVHYKPTKLPQEKPMIPKFKEGVEIIPRDYQIESYYQSTQ
                     DNTLLVLPTGMGKTLVSIMTLLEMFSINDQDNSCGDSKRIALFLVDRVPLVTQQAGAI
                     EAITNLKVCKLYGEINDSRTRAFVRSKEYDVLVSTVGSLVNLLEVRHLNILDFYFITF
                     DEVHHATGEHDFNKVVDYIRKTDLNYRPRILGLTASLVSIGNSTIDIVQRSIKDMEER
                     MLSRVFKPTSLLTNSTQSELQPELVSFKTSGQESLIETSICEFLVSNSKELQIFLNYS
                     DVDLDGVRGNPSFLKALERLDKYSKQLESIYIEYTKVLIELYALLSVLSTEGPKQVLT
                     RLDLLMQSTNQENELYNKLLELRYYISGQYEDRESNFEKGSTRYRKLISSLEYAISDA
                     SDGELQKDLRILVFVETRFGASNLTSMLKKEPFQEYLHTKRLVGHNGVDGMDSEKQQS
```

```
                  IIRKFRDGKCRLIVTTNVLEEGIDVQDCNIVICYDGILSLKSLIQRRGRARSKNESKF
                  IIIYNDDKDLKILDILSSENLLNNSINQILSNRERQAITKNQFLNWKKLNINNEYFTN
                  NQLDGGPLDKDIKCVILSIYHIEIDIETFRELLLFQLPDIINQVDQVSIRMEGDDEYL
                  SSYIIVILNTCIDKEQSIYRLKSLLNINGLKYWMIDRSSFHFEDDNINDLSQSKNKNY
                  SIRSFEYGNMSSPTKYHGIGQLDDSFELFYGKREMFFTKKTRFQRYRLNLKLTDIEST
                  FLLVNDKYQNIVHILLIVKRPLIFEEEKKTHIVFWNKSKPKKDDFGQAFVYRLTLSTI
                  DNEIKEVIKPLIANGFRMYESSLELNCFIQGGINKNNNCMIIDLNLQYIKRNLEIHYL
                  LQCLQSQSSYGRVLNQDFIVEINNLLAFGKQEEAEFIITSLISKRTKFINYKEFIGDC
                  LKNFLPDPYYAKKKDLSTKSGEFTMIRDVFITPSRTIFIQPTLQRSCRAVRKFGSSNF
                  IVVKIVNESLEPLQGVNNKHPLHENRIKPILEGGILVGGEIYSYAGNSNSQLREYSSW
                  FVSNQIGTHTVKIWSGIEHVDNVRKFFRCIGLMFSTTIPTVTLPQNRIYRIQDITRNT
                  HEFTEGCGEIGPELAKHLNENYNFRPSTCAYQVRIGGNKGVLVVNNQAPDPSGIYIRP
                  SMVKFNPIDCGDEHRTLEICSVSTTSRCKLNRQVISLLSTLGTQDNVFFALQDHYLNQ
                  VAQIVNDTNASKQAIVEFFPDITEGELYQDPYIRRILISLYKLKMERIQQKCHIEIKD
                  SRMLLGVCDPTNSLPPNTVFVQLEEEDEDDDDDGRKYEKVIEGLVMVIKNPCTHPGDV
                  RYLKAVDNIRLRHLRNVLVFSTKGDVPNFKEISGSDLDGDRYFFCYDKSLIGNRSESE
                  TAYLVVETVSNNDKKANVFNDPFALSSMYSTNAERQELGKLYHSHLAISDQYGANHKY
                  SIQISKECFKEIDYPKTGIHGTIPKEVNIRLQTVGYPHYMQRENSTRVYYQSKKIMGK
                  MYDQIDQLVYIGDFLPNISLDKSNLVDGYEIYLNSAKILYSQYKLQVHSLLRHYSAES
                  EESIMIGFLDQGFISDKVSKDIKGEMKNDYIKIQQTFENEFLKEFGEQHKENCLLIHR
                  VNIEKKVSAWYHVAYSDLKDDRALGFYWIARSFKQMTLDQESQKKKNILVNSILDHMN
                  KKKDTLFPLYESRLLIVKELTSFLNTLNDTVFKESTLEIIGSTATMLFDENSNLNLYL
                  KLENQNDHMNLAEKNEILNQLKDEIKTSSFISAEACSIIKIQGDIPLFKFKIDKIQCL
                  IGLETSKWYETLVQQQYITKNPSILSILFSIIDWSINSLFVDSETGEIINNSKVYSSL
                  NLQKVIVENSMIPSIFNNKNHNNTTIIIIIIIIIIIIIII"
     exon         <1. .1572
                  /gene="RrpA"
                  /number=1
     intron       1573. .1654
                  /gene="RrpA"
                  /number=1
     exon         1655. .>5731
                  /gene="RrpA"
                  /number=2
ORIGIN

Alignment Scores:
Pred. No.:              6.17e-12        Length:         5731
Score:                  260.50          Matches:        133
Percent Similarity:     39.23%          Conservative:   100
Best Local Similarity:  22.39%          Mismatches:     224
Query Match:            8.65%           Indels:         137
DB:                     3               Gaps:           26

US-09-913-878A-2_COPY_710_1282 (1-573) x DDI314909 (1-5731)

Qy        1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
Db     3362 GGATGTGGTGAAATTGGTCCAGAACTTGCTAAACATTTAAATGAAAATTATAATTTCAGA 3421

Qy       21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
Db     3422 CCATCAACATGTGCATATCAAGTTAGAATAGGTGGAAATAAAGGTGTGTTGGTGGTAAAC 3481

Qy       41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
Db     3482 ---AATCAAGCACCCGATCCATCTGGTATTTATATTAGACCAAGTATGGTGAAATTCAAT 3538

Qy       61 ---CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGlu 79
Db     3539 CCAATCGATTGTGGTGATGAACAT---AGAACATTAGAGATTTGCAGTGTT---TCGACA 3592

Qy       80 LeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAsp---ArgAlaArg 98
Db     3593 ACTTCAAGATGTAAATTAAATAGACAAGTCATCAGTTTACTCTCAACATTGGGTACACAA 3652

Qy       99 AspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGln 118
Db     3653 GATAATGTT------------------------------------------------TTC 3664

Qy      119 PheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGlu 138
Db     3665 TTTGCATTACAAGATCATTATCTAAAT---------CAAGTGGCTCAAATCGTAAATGAT 3715

Qy      139 SerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeu 158
Db     3716 ACCAATGCTTCCAAACAAGCAATCGTTGAATTC-----------------------TTT 3751

Qy      159 ProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLys 178
Db     3752 CCAGATATTACAGAGGGTGAACTT-----------------------------TATCAA 3781

Qy      179 GlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLys 198
Db     3782 GATCCTTACATTCGTAGAATTTTAATTAGTCTTTATAAATTAAAAATGGAGAGAATTCAA 3841

Qy      199 SerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrp 218
Db     3842 CAAAAATGTCATATAGAGATTAAAGATAGTAGAATGTTGTTGGGTGTTTGTGATCCAACA 3901

Qy      219 GlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGlu 238
Db     3902 AATTCATTACCACCAAATACTGTATTTGTTCAATTGGAAGAGGAAGATGAAGATGATGAT 3961

Qy      239 Glu---------SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerPro 255
Db     3962 GATGATGGTCGTAAATATGAAAAAGTAATCGAAGGATTGGTAATGGTAATTAAAAATCCT 4021

Qy      256 AlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSer 275
Db     4022 TGTACTCATCCTGGTGATGTAAGATATCTAAAGGCTGTTGATAATATAAGATTAAGACAT 4081

Qy      276 LeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSer 295
Db     4082 TTAAGAAATGTATTGGTTTTCTCAACAAAAGGTGACGTACCAAACTTTAAAGAAATTTCA 4141

Qy      296 GlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGly 315
Db     4142 GGTTCAGATTTAGATGGTGATCGTTATTTCTTTTGTTACGATAAATCATTGATTGGTAAT 4201

Qy      316 PheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu------------ 331
Db     4202 CGTTCAGAATCAGAG---------------------ACAGCCTATTTAGTTGTCGAGACC 4240

Qy      332 ------------LysLysAspLysThrThrPheLysGln---------LeuMetAlaSer 344
Db     4241 GTTTCAAACAATGATAAAAAGGCCAATGTTTTCAATGATCCATTTGCTTTATCCTCAATG 4300

Qy      345 HisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPhe 364
Db     4301 TACTCTACAAATGCTGAACGTCAAGAA------------------CTTGGTAAATTGTAT 4342

Qy      365 HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCys 384
Db     4343 CAT---------------TCACATTTGGCAATCTCTGATCAATAC--------------- 4372

Qy      385 TyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
Db     4373 ------------GGTGCAAATCACAAATATTCAATTCAAATCTCTAAAGAATGTTTTAAA 4420

Qy      405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
Db     4421 GAAATTGATTATCCAAAAACTGGTATTCATGGCACTATTCCAAAAGAAGTAAACATTCGG 4480

Qy      425 ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMet------------------ 438
Db     4481 TTACAAACTGTAGGT------------TATCCACATTATATGCAAAGAGAGAATTCAACC 4528

Qy      439 ---------TyrLysSerAspSerTrpLeuGlyArg---------------GlyGluPro 450
Db     4529 AGAGTCTATTATCAATCGAAAAAAATTATGGGTAAAATGTATGATCAAATTGATCAGTTG 4588

Qy      451 ThrHisIleIleAspTyrLeu---LysPheSerIleAlaArgProAlaIleAspLysGlu 469
Db     4589 GTTTATATTGGTGATTTCTTACCAAATATTAGTCTTGATAAAAGTAATTTGGTTGATGGC 4648
```

```
Qy        470 LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPhe 489
                 |||   :::   |||:::   |||
Db       4649 TATGAAATCTATTTAAATAGTGCTAAAATC------------------------------ 4678

Qy        490 TrpAspProAspLeuAlaSerTyrTyr---------------ThrPhePheLysGluIle 504
                          |||   |||   |||               :::      :::
Db       4679 ------------TTATATAGTCAGTATAAACTTCAAGTTCATTCATTGTTAAGACATTAT 4726

Qy        505 SerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGlu 524
              |||   :::|||      |||   :::                        ||||||
Db       4727 TCAGCAGAATCTGAAGAATCTATAATG------------------------ATTGGTTTC 4762

Qy        525 ValGluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAsp------------ 540
              ::::::::   :::            :::   :::|||      :::|||
Db       4763 CTTGATCAAGGTTTC---------ATTAGTGATAAAGTTTCAAAAGATATCAAAGGTGAA 4813

Qy        541 SerLysAspProTyrProValArgValAsnGlnValTyrGlu 554
                 |||:::   |||   :::::::::   |||   :::|||
Db       4814 ATGAAAAATGATTAC---ATTAAAATTCAACAAACTTTTGAA 4852

RESULT 12
AP004357/c
LOCUS       AP004357          165701 bp    DNA     linear   PLN 27-NOV-2003
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
            BAC clone:B1074C08.
ACCESSION   AP004357 BA000010
VERSION     AP004357.4  GI:21328130
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T., Yamamoto,K., Sakata,K., Baba,T.,
            Katayose,Y., Wu,J., Niimura,Y., Cheng,Z., Nagamura,Y.,
            Antonio,B.A., Kanamori,H., Hosokawa,S., Masukawa,M., Arikawa,K.,
            Chiden,Y., Hayashi,M., Okamoto,M., Ando,T., Aoki,H., Arita,K.,
            Hamada,M., Harada,C., Hijishita,S., Honda,M., Ichikawa,Y.,
            Idonuma,A., Iijima,M., Ikeda,M., Ikeno,M., Itoh,S., Itoh,T.,
            Itoh,Y., Itoh,Y., Iwabuchi,A., Kamiya,K., Karasawa,W., Katagiri,S.,
            Kikuta,A., Kobayashi,N., Kono,I., Machita,K., Maehara,T.,
            Mizuno,H., Mizubayashi,T., Mukai,Y., Nagasaki,H., Nakashima,M.,
            Nakama,Y., Nakamichi,Y., Nakamura,M., Namiki,N., Negishi,M.,
            Ohta,I., Ono,N., Saji,S., Sakai,K., Shibata,M., Shimokawa,T.,
            Shomura,A., Song,J., Takazaki,Y., Terasawa,K., Tsuji,K., Waki,K.,
            Yamagata,H., Yamane,H., Yoshiki,S., Yoshihara,R., Yukawa,K.,
            Zhong,H., Iwama,H., Endo,T., Ito,H., Hahn,J.H., Kim,H.I., Eun,M.Y.,
            Yano,M., Jiang,J. and Gojobori,T.
  TITLE     The genome sequence and structure of rice chromosome 1
  JOURNAL   Nature 420 (6913), 312-316 (2002)
  MEDLINE   22337376
   PUBMED   12447438
REFERENCE   2  (bases 1 to 165701)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Jun 6, 2002 this sequence version replaced gi:20975426.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'. A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from M13rev to -21M13 of the BAC
            clone. This sequence of B1074C08 clone has an overlap with B1147B04
            clone (DDBJ:AP004641) at the position 1 to 108,335 of 5' end and an
            overlap with B1129H01 clone (DDBJ: AP003370) at the position
            115,348 to 165,701 of 3' end. Detailed information on overlap and
            assembly quality together with annotation of this entry is
            available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1. .165701
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="1"
                     /clone="B1074C08"
     gene            complement(join(1956. .2373,2716. .2783))
                     /gene="B1074C08.1"
     CDS             complement(join(1956. .2373,2716. .2783))
                     /gene="B1074C08.1"
                     /note="contains EST D22801(C1352)
                     unknown protein"
                     /codon_start=1
                     /protein_id="BAC00710.1"
                     /db_xref="GI:21328131"
                     /translation="MDSTLTSQKATATATFIITSTVRINRFLGRTSGATKKPKGASFQ
                     AMESYVLTWFIVFAFDSSDMVIVQCTHSRTLELKSYFCYCYCYLYHFFLLASMYKRPT
                     PFWHSTQPKKNQNFPRRSDPTNEATNRPTEEDRGMDLEDRVGGDQSCQRRSMTTQRWM
                     M"
     gene            join(7311. .7428,7640. .7776)
                     /gene="B1074C08.2"
     CDS             join(7311. .7428,7640. .7776)
                     /gene="B1074C08.2"
                     /note="contains EST AU069673(C0828)
                     unknown protein"
                     /codon_start=1
                     /protein_id="BAC00711.1"
                     /db_xref="GI:21328132"
                     /translation="MHCHCVRCNVRSGYAQQQPFVRACVRAAQPRWKVHSNTRPAARM
                     KLIGSGSGSGGQRRGIARALKEHKARLYIIRRCVVMLLRC"
     gene            join(9087. .9169,9577. .9836,9998. .10158,10204. .10298,
                     10342. .10570)
                     /gene="B1074C08.3"
     CDS             join(9087. .9169,9577. .9836,9998. .10158,10204. .10298,
                     10342. .10570)
                     /gene="B1074C08.3"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAC00712.1"
                     /db_xref="GI:21328133"
                     /translation="MARQQCLDRVGRDGDGAANAVEKMREREVTDMWGPPVGPFFHLP
                     PPLPKPATTAPTPLAYPAPPRLHSRTPAPSPRPRRLPRPSLPAHARAPKRAGFDLNPP
                     PTLSLSLSIPHVGPPPSRAVIAFGFACSRAISSTSPLSKSAIGGRHSPCAGEPFHRRR
                     RRPWFSLTVVLELVRSLPPFPGRSPSLFAGVFPHAGRIPRAEPRRRPSRHGLRPTLPP
                     PAPPVTTRAAAASHGRLVGFEAEPSPAAAPRAVRSARGRSGQSSCGPAVDHPRGPGP"
     gene            join(16765. .17574,19784. .20080)
                     /gene="B1074C08.4"
     CDS             join(16765. .17574,19784. .20080)
                     /gene="B1074C08.4"
                     /codon_start=1
                     /product="putative DNA damage repair protein"
                     /protein_id="BAC00713.1"
                     /db_xref="GI:21328134"
                     /translation="MLGGLYGDLPPPSSSADDDKPSAAGWSSAAKMAPPTLRKPPATF
                     APPPSVLRNSRPAPKAPAAQPPPPPTLPIETTTSTSFQPALVAVQSTVMEEYDPARPN
                     DYEDYRKDKLKRAKEAEVRKELERRREEEERERERELRERERGRDALNISGEEAWKRR
                     AAMSGSAAPRPSSSPPHGDGFAIGNSSSSGLGLGAGGQMTAAQRMMARMGWKEGQGLG
                     KQEQGITAPLVAKKTDRRGGVIVDENSSKQEKKPKSVNFDGPPTRVLLLRNMVGPGEV
                     DDELEEEVASECSKYGTVLRVLIFEITQADFPAEEAVRIFILFERAEEATKAMIDLEG
```

```
                     RFFGGRVVRATFFDEERFGKNQLAPMPGEVAGFD"
     gene            join(22181. .22363,22471. .22640,22731. .22803,22919. .22993,
                     23119. .23229,23337. .23382,23475. .23538,24200. .24302,
                     24389. .24484,24576. .24605,24824. .24923,25151. .25235,
                     25355. .25478)
                     /gene="B1074C08.5"
     CDS             join(22181. .22363,22471. .22640,22731. .22803,22919. .22993,
                     23119. .23229,23337. .23382,23475. .23538,24200. .24302,
                     24389. .24484,24576. .24605,24824. .24923,25151. .25235,
                     25355. .25478)
                     /gene="B1074C08.5"
                     /note="contains ESTs
                     C26310(C12073),D15391(C0569),AU100694(C12073)
                     unknown protein"
                     /codon_start=1
                     /protein_id="BAC00714.1"
                     /db_xref="GI:21328135"
                     /translation="MAPGTTLAPKRRKAEASPSPSPSPMGDSSDGGYSDSDLHDAEES
                     FYSARSGSEDDRQVSSSNDDDDSEEEEQEEREMDEEEDEEDDDDEEMNEEDDEDEGEE
                     MNELEKEYRTLQTNQQNILETLKQHRDDDVSKGQAVKNQKVLWDKALEMRFLLQKAFS
                     TSNKLPKEPIRSMFCDHNQEIEQAYLDLLNSSKQTLGSMMELQEALLERNRATKDVTD
                     TDNSSELNGEDDEWSEVQKLQKRITPFRNSEIDKWQRKTQVTTGAAALKGKLHAFNQN
                     ISDQVTSYMRDPSRMINRMHLRKSTLGVFGEEVGEHENNKEENNTEGDPELVDDSEFY
                     QQLLKEFLESCDAGASESAFYALKKQQHKKRKLVDRRASKSRKIRYHVHEKIANFMAP
                     VPMVIPPMAPKLFENLFGMGNQKSTTA"
     gene            join(26433. .26525,30039. .30091,31996. .32168,32606. .32724)
                     /gene="B1074C08.6"
     CDS             join(26433. .26525,30039. .30091,31996. .32168,32606. .32724)
                     /gene="B1074C08.6"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAC00715.1"
                     /db_xref="GI:21328136"
                     /translation="MFIHRTKLLDAAAQLAVISDRRVFVYTCLMKPSSSSFVVDTISP
                     LSQHRGVTPLLLAVRSRGLAHFFFHRDQPGSLLTPAKSAVMWLTPDVQQATHIAHLTQ
                     LRSVQSTTTEKLIGAGWEPALGAGFPNRHHGGDTDDFQSRCRF"
     gene            36236. .36550
                     /gene="B1074C08.7"
     CDS             36236. .36550
                     /gene="B1074C08.7"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAC00716.1"
                     /db_xref="GI:21328137"
                     /translation="MAAALGQIWRRRGAAAAFPPATTRRRCRRLPSRYHPPPPPPPPL
                     LPGCRYRSYLVEGRRCRRRHRLPSCHHPPSPSPPPLLPGRHSRPDLVEDAASTAGPPP
                     RV"
     gene            complement(38211. .39233)
                     /gene="B1074C08.8"
     CDS             complement(38211. .39233)
                     /gene="B1074C08.8"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAC00717.1"
                     /db_xref="GI:21328138"
                     /translation="MDWASLPADLLFSISSHLREPEDFVRFRAVCPQWRAAVSHKEHA
                     FFQPWIMASRWLEDEYSENVLFYSLSTLKTIKVHVPDMKSRRIAASGSSHLIAIDKDD
                     DLSAVLINPLSGKTTALPRLPKFFHDNGAHGWITGEGVITVVLNNWMSENMALWYRGG
                     GITMKGWAVVPRWKLRLRVSHYLRKLAAYGDQMEMHLIDLGGDNEDSVVLLQETKKVE
                     LLGGCWPRSDEVFKATRPCHHEWFSLYRNVEQEEIPVHDIGNVMVVQSRDSCTRTYMI
                     PASRDFAALGSRNAFYYLWKQFDAGGSYNALFKKCLASEVLTFVKRLPEDWKLSDEWF
                     MPSLKY"
     LTR             42943. .43587
                     /note="5' LTR"
     gene            join(44015. .44093,44667. .44836)
                     /gene="B1074C08.9"
     CDS             join(44015. .44093,44667. .44836)
                     /gene="B1074C08.9"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAC00718.1"
                     /db_xref="GI:21328139"
                     /translation="MPPRRCLEAAAPRRVPRRASLPCAAAEPEGVADLSEGEDPQDYE
```

```
                     QGKSSLFFEHVDPNCEIILFTNKIACNDEHPTCDYVMP"
     gene            join(47189. .47443,47555. .47728)
                     /gene="B1074C08.10"

Alignment Scores:
Pred. No.:                7.07e-10       Length:          165701
Score:                    259.00         Matches:         104
Percent Similarity:       39.34%         Conservative:    75
Best Local Similarity:    22.86%         Mismatches:      162
Query Match:              8.60%          Indels:          114
DB:                       8              Gaps:            17

US-09-913-878A-2_COPY_710_1282 (1-573) x AP004357 (1-165701)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
              |||:::|||:::::::::        :::|||      :::   :::    |||   |||
Db     112198 GGAATTGGCAAAATTACCCCAGATCTTGCATTGGAAGTTGCCGAGAGGCTGCAGCTGACA 112139

Qy         21 Asp---ValProSerAlaValGlnGlyArgPheGlySerAlaLysGly---------Met 36
              |||      |||||||||   |||   ||||||         ||||||         :::
Db     112138 GATAACCCGCCATCTGCTTATCAGATCAGGTTTGCTGGCTTCAAGGGTGTCATAGCTGTC 112079

Qy         37 TrpValIleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln 56
              |||                     ||||||      :::            ||||||
Db     112078 TGG------------CAAGGACATGGTGATGGGACACGGCTTTTCCTGAGGCCAAGCATG 112031

Qy         57 ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerVal 76
              ||||||:::|||   :::   :::                   |||||||||   |||
Db     112030 AGGAAGTTTGAGTCTAACCATTTG----------------GTGTTAGAGGTGGTCTCC--- 111989

Qy         77 AlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArg 96
                 ::::::   :::              ||||||   |||::::::   :::|||
Db     111988 TGGACAAAGTTCCAGCCAGGATTCTTAAATCGACAGATTATAATATTGCTATCCTCA--- 111932

Qy         97 AlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGln 116
                                          :::   |||   :::         :::|||
Db     111931 -----------------------CTGAATGTCCCAGATTCTATCTTTTGGCAAATGCAA 111896

Qy        117 ArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpVal 136
                                        ||||||      :::      |||
Db     111895 GAGACCATGCTTTCTAAC---------CTCAACAATATTCTATCAGACAGAGATGTTGCT 111845

Qy        137 TyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAla 156
              :::|||      ::::::
Db     111844 TTTGAGGTTTTAACAACTTCATGTGCT--------------------------------- 111818

Qy        157 GlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspPro 176
                             |||         ::::::::::::||||||:::|||
Db     111817 ------------GATGATGGAAACACTGCAGCATTGATGCTCAGTGCTGGCTTTGAACCT 111770

Qy        177 LysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThr 196
              :::   :::   :::|||:::   :::         ::::::      :::
Db     111769 AGAACTGAACCACACTTGAAAGCAATGCTCTTGGCTATAAGGTCTGCACAATTGCAGGAT 111710

Qy        197 LeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAsp 216
              |||      |||      |||   |||   :::      ::::::         |||
Db     111709 CTTTTGGAAAAAGCAAGGATATTTGTGCCAAAGGGAAGGTGGTTGATGGGCTGTCTTGAT 111650

Qy        217 PheTrpGlyValLeuGluGlu-----------------------------------Asn 224
                    |||||||||||:::                                   |||
Db     111649 GAGCTTGGGGTTCTTGAGCAAGGGCAGTGCTTTATTCGGGCAACAGTTCCATCATTGAAT 111590

Qy        225 GluValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeu 244
                       |||         |||:::|||            :::::::::      ::::::
Db     111589 AGTTATTTTGTTAAGCATGGGTCAAGATTTTCATCAACAGATAAAAACACAGAGGTCATT 111530

Qy        245 SerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArg 264
                       |||::::::|||::::::|||               |||   |||::::::
Db     111529 TTGGGT---ACTGTGGTAATAGCAAAGAATCCCTGTCTTCATCCAGGGGATGTCCGCATC 111473

Qy        265 ValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThr 284
              :::   ||||||      ||||||||||||   |||   |||      ::::::|||
Db     111472 CTTGAAGCAGTTGATGTGCCCGAACTGCATCATCTGGTTGATTGTTTGGTGTTCCCCCAG 111413
```

```
Qy        285 LysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAla 304
              ||||||:::   |||   |||   :::   ||||||  |||   |||||||||:::
Db     111412 AAAGGTGAGAGGCCACATGCTAACGAGGCATCTGGGAGCGATCTTGATGGGGATCTCTAC 111353

Qy        305 TrpValCysTrpAspProGluIleValAsp------------------------------ 314
              :::|||   ||||||   :::::::::
Db     111352 TTTGTGACATGGGATGAGAAACTTATACCTCCAGGCAAGAAGAGCTGGAACCCTATGGAC 111293

Qy        315 ---------------------------------------------------GlyPheVal 317
                                                                 :::
Db     111292 TACTCCCCACCTGAAGCAAAACAACTTCCGCGCCAAGTATCTCAACATGTAAGTTCAATA 111233

Qy        318 AsnAlaGluMetPro------------------LeuGluProAspLeuSerArgTyrLeu 331
                 ||||||                       |||   |||            :::|||
Db     111232 ACTCTTGAGATGTCAGCAATGTCCCATAATTTTTTGTTTCCTACATACATGTCCTTTTTA 111173

Qy        332 LysLysAspLysThrThrPhe--------------------------------------- 338
                                 :::|||
Db     111172 TACGATGGATAACCCTCCTTTCATTGGAAGCATATATGGGTTTTCTTTACAGAACTGAGA 111113

Qy        339 -----------LysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGln 354
                         |||   |||      |||   ||||||:::   ||||||
Db     111112 TAAAGAAATTAAACGCAAGCTATGGTCGGGCAC---ACTGGCACACCAGCAAAAACAGGG 111056

Qy        355 ThrThrTyr---------AspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 371
              ::::::            |||::::::|||:::|||            |||   ||||||
Db     111055 AATAGCTTTAGAATCTCAGATATATTACAGCAATCA---------AAATTAAGTCCAAAC 111005

Qy        372 PheLeuGly---------MetCysThrAsnTyrLysGluArgLeu 383
                    |||         :::|||      |||:::      |||
Db     111004 CAGAGAGGTCCTCCTAGCATTTGTAAACCTTATAGGCATCATCTC 110960
```

```
RESULT 13
AX204840
LOCUS       AX204840                3591 bp    DNA     linear   PAT 30-AUG-2001
DEFINITION  Sequence 2 from Patent WO0155407.
ACCESSION   AX204840
VERSION     AX204840.1  GI:15394184
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Beclin,C., Elmayan,T., Mourrain,P. and Vaucheret,H.
  TITLE     Novel sgs2 plant gene and use thereof
  JOURNAL   Patent: WO 0155407-A 2 02-AUG-2001;
            Rhobio (FR) ; INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (FR)
FEATURES             Location/Qualifiers
     source          1..3591
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
     CDS             1..3591
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC60174.1"
                     /db_xref="GI:15394185"
                     /db_xref="REMTREMBL:CAC60174"
                     /translation="MGSEGNMKKSVVTQVSIGGFGESTTAKQLTDYLEDEVGIVWRCR
                     LKTSWTPPGSYPNFEIADTSNIPSIDEYKKVEPHAFVHFAVFESAGRAMDAAGQCNLI
                     LDGQPLKVSLGPKNPYSLNQRRRTTVPYKLAGITLEIGTLVSRDDFFVSWRAEGVDFL
                     VDPFDNTCKFCFRKSTAFSFKDAVMHAVINCDYKLELLVRDIQTVRQYKTLHGFVLIL
                     QLASSPRVWYRTADDDIYDTVPGDLLDDDDPWIRTTDFTQVGAIGRCHSYRVLISPRY
                     ENKLRTALDYFRMRRVQEERVRWPPRIRNEPCFGEPVSDHFFCIHHKEGISFEIMFLV
                     NSVLHRGVFNQFQLTERFFDLLRNQPKDVNIASLKHLCTYKRPVFDAYKRLKLVQEWI
                     QKNPKLLGSHEQSEDISEIRRLVITPTRAYCLPPEVELSNRVLRRYKAVAERFLRVTF
                     MDESMQTINSNVLSYFVAPIVKDLTSSSFSQKTYVFKRVKSILTDGFKLCGRKYSFLA
                     FSANQLRDRSAWFFAEDGKTRVSDIKTWMGKFKDKNVAKCAARMGLCFSSTYATVDVM
                     PHEVDTEVPDIERNGYVFSDGIGTITPDLADEVMEKLKLDVHYSPCAYQIRYAGFKGV
                     VARWPSKSDGIRLALRDSMKKFFSKHTILEICSWTRFQPGFLNRQIITLLSVLGVPDE
                     IFWDMQESMLYKLNRILDDTDVAFEVLTASCAEQGNTAAIMLSAGFKPKTEPHLRGML
                     SSVRIAQLWGLREKSRIFVTSGRWLMGCLDEAGILEHGQCFIQVSKPSIENCFSKHGS
                     RFKETKTDLEVVKGYVAIAKNPCLHPGDVRILEAVDVPQLHHMYDCLIFPQKGDRPHT
                     NEASGSDLDGDLYFVAWDQKLIPPNRKSYPAMHYDAAEEKSLGRAVNHQDIIDFFARN
                     LANEQLGTICNAHVVHADRSEYGAMDEECLLLAELAATAVDFPKTGKIVSMPFHLKPK
                     LYPDFMGKEDYQTYKSNKILGRLYRRVKEVYDEDAEASSEESTDPSAIPYDAVLEIPG
                     FEDLIPEAWGHKCLYDGQLIGLLGQYKVQKEEEIVTGHIWSMPKYTSKKQGELKERLK
                     HSYNSLKKEFRKVFEETIPDHENLSEEEKNILYEKKASAWYHVTYHPEWVKKSLELQD
                     PDESSHAAMLSFAWIAADYLARIKIRSREMGSIDSAKPVDSLAKFLAQRL"
ORIGIN

Alignment Scores:
Pred. No.:              5.43e-12        Length:         3591
Score:                  258.00          Matches:        140
Percent Similarity:     37.58%          Conservative:   99
Best Local Similarity:  22.01%          Mismatches:     247
Query Match:            8.57%           Indels:         150
DB:                     6               Gaps:           28

US-09-913-878A-2_COPY_710_1282 (1-573) x AX204840 (1-3591)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
                |||:::|||   ::::::      :::|||      :::   :::   |||   |||
Db       1759 GGAATTGGTACAATCACACCTGACCTCGCTGACGAAGTAATGGAGAAACTTAAGTTG--- 1815

Qy         21 AspVal---------ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrp 37
              ||||||                  |||   |||   |||   |||:::   |||||||:::
Db       1816 GATGTGCACTACAGCCCTTGTGCTTATCAGATACGTTACGCAGGTTTCAAAGGGGTT--- 1872

Qy         38 ValIleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln--- 56
                                                   :::      :::||||||:::
Db       1873 ------------------------------------GTTGCTCGTTGGCCATCAAAAAGT 1896

Qy         57 --------------------ArgLysTrpGluCysAspPheValAspLysHisGlnArg 69
                                  |||               |||      ||||||
Db       1897 GATGGAATCAGGCTAGCCCTTCGAGACAGTATGAAGAAGTTCTTTTCCAAACAT---ACG 1953

Qy         70 ThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeu 89
                |||||||:::   |||      :::      :::      ||||||   |||:::
Db       1954 ATCTTGGAGATCTGTTCC---TGGACGAGGTTTCAACCTGGGTTCTTAAATCGGCAGATA 2010

Qy         90 LeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAsp 109
              :::   :::|||                                    :::   |||
Db       2011 ATTACCCTTCTATCC--------------------------GTACTAGGTGTTCCGGAT 2043

Qy        110 ArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgPro 129
                :::      |||:::|||                  :::   |||||||||
Db       2044 GAAATATTCTGGGATATGCAG---------GAATCCATGCTCTATAAACTGAACCGCATC 2094

Qy        130 ValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHis 149
              ::::::            :::|||   ::::::
Db       2095 CTTGATGATACAGATGTGGCATTTGAAGTTCTCACGGCATCATGTGCT------------ 2142

Qy        150 GlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeu 169
                                                |||      |||         :::
Db       2143 --------------------------------GAACAGGGAAACACTGCAGCTATCATG 2169

Qy        170 MetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeu 189
              ::::::::||||||   ||||||   :::   :::|||:::   :::         :::
Db       2170 CTTAGTGCAGGTTTCAAACCAAAAACCGAGCCGCATCTACGCGGGATGTTGTCTTCAGTC 2229

Qy        190 GlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAla 209
              :::      :::         |||:::   |||      |||   |||
Db       2230 AGAATTGCACAACTCTGGGGTCTCAGAGAAAAATCTCGTATTTTTGTTACTTCAGGAAGG 2289

Qy        210 TyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGly 229
              ::::::         |||      |||:::||||||      :::      :::
Db       2290 TGGCTAATGGGTTGCCTAGACGAAGCAGGGATACTTGAACATGGCCAATGCTTTATTCAA 2349

Qy        230 PheSer-----------------------------------SerLysPheArgAspGlu 237
                 |||                                    |||:::|||::::::
Db       2350 GTCTCTAAACCGTCTATAGAAAATTGTTTCTCCAAACATGGGTCTCGTTTTAAGGAGACA 2409
```

```
Qy        238 GluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHis 257
Db       2410 AAGACAGATCTGGAAGTAGTTAAAGGCTAT---GTAGCCATTGCTAAGAATCCTTGTCTT 2466

Qy        258 PheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLys 277
Db       2467 CACCCAGGGGATGTAAGGATTTAGAAGCTGTTGATGTACCCCAGCTGCATCACATGTAT 2526

Qy        278 AspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGly 297
Db       2527 GACTGCCTTATTTTCCCTCAGAAAGGTGATAGGCCGCATACAAACGAAGCTTCTGGCAGT 2586

Qy        298 AspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheVal 317
Db       2587 GACCTTGACGGGGACCTGTACTTTGTGGCTTGGGATCAGAAACTCATC------------ 2634

Qy        318 AsnAlaGluMetProLeuGluProAspLeuSerArgTyr------LeuLysLysAspLys 335
Db       2635 ------------------CCTCCCAACAGGAAAAGCTATCCGGCCATGCATTATGATGCA 2676

Qy        336 ThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThr 355
Db       2677 GCTGAAGAGAAGAGTTTA------------------GGCCGTGCTGTCAACCACCAGGAC 2718

Qy        356 ThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMet 375
Db       2719 ATAATCGATTTCTTTGCAAGAAAC------------TTGGCGAATGAGCAGTTGGGCACA 2766

Qy        376 CysThrAsn------------TyrLysGluArgLeuCysTyrIleAsnAsnSerValSer 391
Db       2767 ATTTGCAATGCACACGTCGTTCATGCTGATAGAAGTGAGTAT---------GGAGCCATG 2817

Qy        392 AsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGln 411
Db       2818 GACGAAGAATGTTTGCTACTGGCAGAACTAGCTGCCACTGCAGTTGATTTCCCAAAGACA 2877

Qy        412 GlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAla 431
Db       2878 GGGAAAATTGTGTCAATGCCC------TTCCACCTAAAACCAAAACTCTACCCAGATTTC 2931

Qy        432 LeuSerLeuProAsp---ProMetTyrLysSerAspSerTrpLeuGlyArg--------- 447
Db       2932 ATGGGAAAAGAAGACTACCAAACTTACAAGTCGAACAAAATCTTGGGTCGGCTTTACAGA 2991

Qy        448 ---------------------------------------------------GlyGluPro 450
Db       2992 CGGGTAAAAGAGGTTTATGATGAAGATGCAGAAGCTTCCTCAGAAGAAAGCACAGACCCA 3051

Qy        451 ThrHisIle---------------IleAspTyrLeuLysPheSerIleAlaArgPro--- 464
Db       3052 AGTGCCATCCCTTATGACGCTGTTCTTGAAATACCGGGATTTGAAGATTTGATCCCTGAG 3111

Qy        465 ------------------AlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLys 478
Db       3112 GCATGGGGTCACAAATGTTTGTACGACGGGCAACTCATTGGTCTTCTTGGGCAATACAAG 3171

Qy        479 AlaAlaLysAspThrGlu---AspGlyAlaHisPheTrpAspProAspLeuAlaSerTyr 497
Db       3172 GTGCAGAAAGAGGAAGAGATTGTGACGGGTCACATCTGGTCC---------ATGCCCAAA 3222

Qy        498 TyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThr 517
Db       3223 TACACAAGCAAGAAACAAGGCGAACTGAAAGAAAGA------------------------ 3258

Qy        518 LeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGlu 537
Db       3259 CTGAAGCATTCTTATAATTCCCTTAAAAAGGAGTTC---AGGAAAGTATTTGAGGAAACA 3315

Qy        538 MetArgAspSerLysAspProTyrProValArgValAsnGlnValTyrGluLys------ 555
Db       3316 ATCCCTGACCATGAAAATCTCAGCGAGGAGGAAAAAAACATCTTGTATGAGAAGAAGGCT 3375

Qy        556 ------TrpCysAlaIleThr------ProGluAlaMetAspLysSer 567
```

```
Db       3376 TCAGCTTGGTATCATGTAACTTACCATCCCGAGTGGGTGAAGAAGTCT 3423
```

```
RESULT 14
AF443073
LOCUS       AF443073                3524 bp    DNA     linear   PLN 03-DEC-2001
DEFINITION  Phomopsis sp. CMW 5588 putative RNA-dependent RNA polymerase
            (rdp-1) gene, partial cds.
ACCESSION   AF443073
VERSION     AF443073.1  GI:17227101
KEYWORDS    .
SOURCE      Phomopsis sp. CMW 5588
  ORGANISM  Phomopsis sp. CMW 5588
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Diaporthales; Valsaceae; mitosporic Valsaceae;
            Phomopsis.
REFERENCE   1  (bases 1 to 3524)
  AUTHORS   Preisig,O. and Wingfield,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-2001) Forestry and Agricultural Biotechnology
            Institute, University of Pretoria, 72 Lunnon Street, Pretoria 0002,
            South Africa
FEATURES             Location/Qualifiers
     source          1..3524
                     /organism="Phomopsis sp. CMW 5588"
                     /mol_type="genomic DNA"
                     /isolate="CMW 5588"
                     /db_xref="taxon:179291"
                     /country="South Africa"
     gene            <1..>3424
                     /gene="rdp-1"
     mRNA            join(<1..2092,2148..>3424)
                     /gene="rdp-1"
                     /product="putative RNA-dependent RNA polymerase"
     CDS             join(<1..2092,2148..3424)
                     /gene="rdp-1"
                     /note="RDP-1"
                     /codon_start=1
                     /product="putative RNA-dependent RNA polymerase"
                     /protein_id="AAL38011.1"
                     /db_xref="GI:17227102"
                     /translation="HKLKHLPASLVKPVGQSVEFGRWTTYYFDINSSFSLSEWESVRT
                     YLADYNVKTVQVNRFELVTPGPSNAWRLLDNSAEMTLELLSSTNTYHLDFQVRYQLEV
                     CISHGLLDEYSVNAQFLDKLNSFDADRARMMLEGVAEASIQLHEPMKIFDDPKVLHYW
                     PNARVPPYATLVRRAVITPTTIYFKTPCVELTNRILRKYSDLNDRFLRVQFTDEITFG
                     KIFSSQDSNKDDNLYIRVHRVMQNGISIGDRHYRFLAFSNSQFRENGAFFFCETDHVT
                     CDSIREWMGDFRHIRSVGKFAARMGQCFTTTRQVNGISIPNIRKIDDIERQTGDHTWN
                     FTDGVGKISNFFAKLIASERDLPETPSCFQMRIGGGKGVLVVWPDVPPSEVHVRPSQE
                     KFKAVYNVLEIIKTFAFSHATFNKQVIPVLVALGVDNSVFVSMLDDELREYDEALADS
                     MKAGELLRSQVDENQTTLTMAEMVDTFMDSKEPFLWTLLRLWKCWILKRLKHKFAISV
                     KKSAMVFGVVDEIGVLRGHSQATEGRGYNSIESLPQIFLQVPIEGSDGRSTTNYEVIT
                     GICVVGRNPSLHPGDVRVVEAVDVPELRHLKNVVVFPKTGDRDIPSMCSGGDMDGDDY
                     FVYWDERLIPTEWDHPPLNHDAGSSTLDKPADVTIEDVTRFFAQYMKNDSLGRIATAH
                     FAQADQLSGGVKNPKCIELAKLHSMAVDYIKSGRPAVMKRHLQPRKWPHWMEKEKKSN
                     YRSYSALGKIYDRIKIEEFHAAYEMPFDARILSRYQLEADTLAKASKIKATYDIAMRR
                     LMGQHEAPVTEFEIWSTFILSRPRVGSDYKLQENVGRELAALKLRFRAECMEAVTGNI
                     QTEGFAPASSVVNLEKLDRFVAAMYTVTHNDVRAALRERSMPKPNGEGGSEEIQMPLI
                     SFPWLFHRELARVALGRGGDVRPLQKSVWLNRGESQQTRDRDVAVEDAEFEIPGDTAL
                     PDAIDLEVRTQQAVGDVQSNKEGYTTERGDSGADGEIGDDCVRTSTGQVVHRGQMLAL
                     FTETGEEVQGQQRSMTPDISPAGTQGLQSADSVSNNSLVADDSTGSSCPASPRSERRE
                     VVEEEEDIEYEEVDGAGYEEVEDALEALARKLGI"
ORIGIN

Alignment Scores:
Pred. No.:              1.16e-11        Length:         3524
Score:                  254.00          Matches:        140
Percent Similarity:     39.02%          Conservative:   84
Best Local Similarity:  24.39%          Mismatches:     216
Query Match:            8.44%           Indels:         134
DB:                     8               Gaps:           25

US-09-913-878A-2_COPY_710_1282 (1-573) x AF443073 (1-3524)
```

```
Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
Db       1012 GGAGTGGGCAAGATATCCAACTTTTTGCAAAACTGATCGCAAGCGAGCGCGATTTGCCA 1071

Qy         21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
Db       1072 GAGACGCCTTCATGCTTCCAGATGAGAATAGGCGGCGGCAAGGGCGTGCTCGTGGTCTGG 1131

Qy         41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
Db       1132 CCTGACGTTCCCCCAAGCGAA------GTTCACGTCCGACCTTCTCAAGAGAAATTCAAG 1185

Qy         61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
Db       1186 GCTGTGTAT--------------AACGTCCTCGAGATCATCAAGACCTTTGCA---TTT 1227

Qy         81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
Db       1228 TCCCACGCCACTTTCAATAAGCAGGTCATACCTGTCCTAGTCGCCCTTGGTGTGGAC--- 1284

Qy        101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
Db       1285 ------------AATTCCGTCTTCGTTAGCATGCTCGATGATGAGCTCAGAGAGTACGAC 1332

Qy        121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
Db       1333 GAA---------GCTTTG-----------------------------------GCAGAT 1347

Qy        141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
Db       1348 TCCATGAAGGCTGGTGAACTCCTTCGATCCCAAGTC--------------------GAT 1386

Qy        161 SerGlnGluGluThrLeuAsnPheLeu---MetAsnSerGlyPheAspProLysLysGln 179
Db       1387 GAAAACCAAACGACATTGACCATGGCAGAGATGGTCGACACCTTCATGGACTCCAAGGAG 1446

Qy        180 LysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArg---LysCys---------Asp 195
Db       1447 CCATTCCTT-----------TGGACCTTGCTTCGCTTGTGGAAGTGTTGGATATTGAAG 1494

Qy        196 ThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAla 215
Db       1495 AGATTGAAGCACAAATTCGCCATCAGTGTCAAGAAAAGCGCCATGGTCTTCGGCGTTGTC 1554

Qy        216 AspPheTrpGlyValLeuGluGluAsnGluValHisVal---------GlyPheSerSer 232
Db       1555 GACGAGATTGGTGTGCTGAGGGGTCACTCGCAAGCCACTGAAGGCAGGGGATACAATAGC 1614

Qy        233 -LysPheArgAspGluGluGluSerPhe------------------------------ 241
Db       1615 ATCGAGTCGCTCCCACAAATCTTCCTTCAAGTGCCTATTGAGGGCAGCGATGGGAGGAGT 1674

Qy        242 -------ThrLeuLeuSerAspCys-AspVal-LeuValAlaArgSerProAlaHisPhe 258
Db       1675 ACCACCAACTATGAAGTCATCACTGGGATTTGCGTCGTTGGCCGCAATCCATCACTCCAT 1734

Qy        259 ProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAsp 278
Db       1735 CCTGGAGATGTACGAGTTGTCGAGGCAGTAGACGTGCCTGAGCTTCGACATCTGAAGAAC 1794

Qy        279 ValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAsp 298
Db       1795 GTCGTCGTCTTCCCAAAGACTGGCGACAGAGACATTCCCAGCATGTGTTCTGGAGGTGAC 1854

Qy        299 TyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleValAsp----------- 314
Db       1855 ATGGACGGGGATGACTACTTTGTCTATTGGGATGAGAGACTCATTCCAACTGAGTGGGAC 1914

Qy        315 -----------------------GlyPheValAsnAlaGluMetProLeuGlu------ 324
Db       1915 CATCCACCCCTGAACCACGATGCGGGATCATCCACCTTGGACAAGCCGGCAGACGTCACA 1974

Qy        325 ---ProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAla 343
Db       1975 ATTGAGGACGTCACCAGATTC--------------------TTTGCACAGTATATGAAG 2013

Qy        344 SerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSer 363
Db       2014 AATGACTCCCTGGGCCGTATCGCC--------------------------------ACT 2040

Qy        364 PheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeu 383
Db       2041 GCGCACTTTGCCCAGGCAGATCAACTGTCAGGGGGCGTCAAGAACCCAAAA--------- 2091

Qy        384 Cys--------------------TyrIleAsn-AsnSerValSerAsnLysProAlaIl 396
Db       2092 TGTAAGCATGCCGCTATCAGCTTATTCTTGATGAAACAATCTGACAGCTCGAATAGGCAT 2151

Qy        396 eIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAs 416
Db       2152 CGAACTCGCCAAGCTCCACTCCATGGCTGTCGACTATATTAAGTCTGGC----------- 2200

Qy        416 nGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAs 436
Db       2201 -------AGACCTGCGGTCATGAAGAGACACCTC-----------------------CA 2229

Qy        436 pProMetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTy 456
Db       2230 ACCTCGA---AAATGGCCGCATTGGATGGAAAAGGAGAAGAAGTCT---------AACTA 2277

Qy        456 rLeuLysPheSerIleAlaArgProAlaIleAspLys---GluLeuGluAlaPheHisAs 475
Db       2278 TCGTTCTTACTCGGCGCTCGGCAAGATCTATGACCGGATCAAGATCGAAGAATTCCATGC 2337

Qy        475 nAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAl 495
Db       2338 GGCATACGAAATGCCCTTTGACGCCCGC-----------------------ATCCTCTC 2373

Qy        495 aSerTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPh 515
Db       2374 GCGCTACCAGCTAGAGGCAGATACCCTGGCCAAGGCCAGCAAAATCAAAGCCACATATGA 2433

Qy        515 eThrThrLeuLysAsnArgIleGlyGluValGlu 526
Db       2434 CATCGCCATGCGCCGCCTCATGGGACAACACGAG 2467
```

RESULT 15
AC105931/c

```
LOCUS       AC105931               132716 bp    DNA     linear   HTG 11-JAN-2002
DEFINITION  Magnaporthe grisea chromosome 7 clone 21J15, *** SEQUENCING IN
            PROGRESS ***, 31 unordered pieces.
ACCESSION   AC105931
VERSION     AC105931.1  GI:18129404
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Magnaporthe grisea
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 132716)
  AUTHORS   Dean,R.A. Dr, Mitchell,T. Dr, Thon,M. Dr, Brown,D.E., Taro,A. and
            Papalas,J.
  TITLE     Magaporthe Grisea chromosome 7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 132716)
  AUTHORS   Dean,R.A. Dr, Mitchell,T. Dr, Thon,M. Dr and Brown,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2002) Plant Pathology - Fungal Genomics
            Laboratory, North carolina State University, 840 Main Campus Drive,
            Raleigh, NC 27606, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
* be preserved.
*        1      981: contig of 981 bp in length
*      982     1081: gap of unknown length
*     1082     1792: contig of 711 bp in length
*     1793     1892: gap of unknown length
*     1893     2493: contig of 601 bp in length
*     2494     2593: gap of unknown length
*     2594     3308: contig of 715 bp in length
*     3309     3408: gap of unknown length
*     3409     4246: contig of 838 bp in length
*     4247     4346: gap of unknown length
*     4347     4612: contig of 266 bp in length
*     4613     4712: gap of unknown length
*     4713     6130: contig of 1418 bp in length
*     6131     6230: gap of unknown length
*     6231     6973: contig of 743 bp in length
*     6974     7073: gap of unknown length
*     7074     8699: contig of 1626 bp in length
*     8700     8799: gap of unknown length
*     8800    10895: contig of 2096 bp in length
*    10896    10995: gap of unknown length
*    10996    12973: contig of 1978 bp in length
*    12974    13073: gap of unknown length
*    13074    15411: contig of 2338 bp in length
*    15412    15511: gap of unknown length
*    15512    19071: contig of 3560 bp in length
*    19072    19171: gap of unknown length
*    19172    21861: contig of 2690 bp in length
*    21862    21961: gap of unknown length
*    21962    24510: contig of 2549 bp in length
*    24511    24610: gap of unknown length
*    24611    27958: contig of 3348 bp in length
*    27959    28058: gap of unknown length
*    28059    32486: contig of 4428 bp in length
*    32487    32586: gap of unknown length
*    32587    36925: contig of 4339 bp in length
*    36926    37025: gap of unknown length
*    37026    43846: contig of 6821 bp in length
*    43847    43946: gap of unknown length
*    43947    48950: contig of 5004 bp in length
*    48951    49050: gap of unknown length
*    49051    54904: contig of 5854 bp in length
*    54905    55004: gap of unknown length
*    55005    60187: contig of 5183 bp in length
*    60188    60287: gap of unknown length
*    60288    65586: contig of 5299 bp in length
*    65587    65686: gap of unknown length
*    65687    71720: contig of 6034 bp in length
*    71721    71820: gap of unknown length
*    71821    80013: contig of 8193 bp in length
*    80014    80113: gap of unknown length
*    80114    88629: contig of 8516 bp in length
*    88630    88729: gap of unknown length
*    88730    96184: contig of 7455 bp in length
*    96185    96284: gap of unknown length
*    96285   102718: contig of 6434 bp in length
*   102719   102818: gap of unknown length
*   102819   112425: contig of 9607 bp in length
*   112426   112525: gap of unknown length
*   112526   120940: contig of 8415 bp in length
*   120941   121040: gap of unknown length
*   121041   132716: contig of 11676 bp in length.
FEATURES             Location/Qualifiers
     source          1..132716
                     /organism="Magnaporthe grisea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:148305"
                     /chromosome="7"
                     /clone="21J15"
ORIGIN

Alignment Scores:
Pred. No.:             1.41e-09       Length:          132716
```

```
Score:                   254.00        Matches:         118
Percent Similarity:      37.83%        Conservative:    56
Best Local Similarity:   25.65%        Mismatches:      158
Query Match:             8.44%         Indels:          128
DB:                      2             Gaps:            21

US-09-913-878A-2_COPY_710_1282 (1-573) x AC105931 (1-132716)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
              ||||||||::::::|||            |||       |||    :::   |||   |||
Db     108509 GGCGTTGGCAAAATCTCTAGCTTGTGGGCTGTTGTCATTTCCGAGAATCTGAAGCTGGAC 108450

Qy         21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
                 :::|||||||||    |||   |||   |||      ||||||:::   |||:::
Db     108449 TATATCCCTTCGGCGTTCCAGTTCCGTATGGGCGGCTGTAAGGGTGTTCTAGTGGTCTGG 108390

Qy         41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
                 |||         |||   |||      |||         ||||||||   |||:::|||
Db     108389 GAAGACGTGAAAGGCCAAGAG------ATTCACATACGTCCTTCTCAGCTCAAGTTTGAG 108336

Qy         61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
                 |||               |||   |||         :::|||         |||
Db     108335 ACGGAC------------CATAGTCGCCTTGAGGTCATTCGC------TGTTCTCACTTC 108294

Qy         81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
                    |||   ||||||   |||   :::   ||||||
Db     108293 TCTGTGGCGTCTCTCAACCGACAAACAATCTCCGTCTTG-------------------- 108255

Qy        101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeu---GlnArgGlnPhe 119
                       ::::::|||   :::         |||::::::   :::
Db     108254 ------------ACTTCCCTGGGAGTCAAA------AATGAGGTGTTTGAAAGTATGTTG 108213

Qy        120 SerGluGlnLysHisAlaLeuAsnArgProValGlu------------------------ 131
              :::||||||            ||||||   :::|||
Db     108212 GCAGAGCAGCTGGCAAAATACAACAGGGCGATGGAAGATGCGACAGCAGCTATGACGCTC 108153

Qy        132 PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArg 151
                 |||:::|||   |||
Db     108152 CTGACACAGTACGTTGACGAA--------------------------------------- 108132

Qy        152 ValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsn 171
                                    :::                  ||||||   |||:::
Db     108131 -----------------------AATCAAATGACCAACACATTGGCCCGGCTTGTGAAG 108096

Qy        172 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIle---------------Ala 186
              :::||||||         :::   ::::::      :::
Db     108095 GCTGGCTTTATGGAGTCGCATGAGCCTTTTGTCTGCACCATGCTGCGCCTATGGAGGACT 108036

Qy        187 TrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGly 206
              |||   ::::::               |||||||||   |||      :::   |||
Db     108035 TGGTCAGTTAAG--------------ACTCTCAAGGAAAAGGCAAGACTCGTCGTTGAG 107991

Qy        207 ArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluVal 226
              :::   |||::::::            |||            |||      :::
Db     107990 AAGGGTGCATTCGTTCTTGGCTGTGTCGACGAAACCAAGACCTTACATGGCCACTTCACC 107931

Qy        227 HisValGlyPheSerSerLysPheArgAspGlu--------------------GluGlu 239
              |||:::         |||         |||
Db     107930 CATATTCCCGCCAGTCCCAACCAGAGGCCCCGCCTCAATGGCAATATGAAGCTGCCTGAG 107871

Qy        240 SerPheThrLeuLeuSerAsp-----------------------------------Cys 247
                 |||      :::   |||                                   |||
Db     107870 ATATTCCTTCAGGTCCCTGATGGCCCTAGCCGTGACTATAAAGTCATAACCGGACTTTGC 107811

Qy        248 AspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAla 267
                    ||||||   |||:::|||:::         |||   ||||||:::   |||:::|||
Db     107810 ------CTTGTTGGCCGCAACCCGTCTTTACACCCTGGAGACATCCGCGTGGTCCAGGCA 107757

Qy        268 ValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAsp 287
              :::      |||:::|||      ::::::||||||::::::|||:::      ||||||
Db     107756 ATTGATGTCCCTCAATTGCGACACATCAGGGACGTCGTTGTTTTGCCCAAACAGGAGAT 107697

Qy        288 ValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCys 307
```

```
                           :::             |||||||||||   ||||||||       :::|||
Db     107696 AGAGACATTCCTGGCATGTGCTCTGGAGGAGACTTGGATGGTGACGATTTCTTCGTCTTG 107637

Qy        308 TrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeu 327
              ||||||      :::                                   :::||||||   |||
Db     107636 TGGGATGAGCGGCTT-----------------------CTCCCTTTA---CCTTCGCAG 107604

Qy        328 SerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThr 347
                    |||                                          ||||||:::
Db     107603 CGCAATTATAAGCCG-----------------------------GCAAGCTATCATCCC 107574

Qy        348 GlySerAlaAla---------LysGluGlnThrThrTyrAspMetIleGlnLysSerPhe 364
                    |||                :::|||   |||      |||:::   |||       |||
Db     107573 CCTCCTGCCCGAGAGCTTGACCGGGAAGTTACGGCAGATGATCTTTGCCAG------TTT 107520

Qy        365 HisPheAlaLeuGlnProAsnPhe--------------------------LeuGlyMet 375
                                 ||||||                           |||
Db     107519 TTTGTGACATACATGAAGAACTTTTCACTTCCAAGGATTGCTCATGCTCACCTGGCGCAG 107460

Qy        376 CysThrAsnTyr------LysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLys 393
                         |||      |||:::     |||
Db     107459 GCAGACAAATATGGCGCAAAAGACGAGAAGTGT-------------------------- 107427

Qy        394 ProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIle 413
                    :::   |||:::   |||            |||||      |||    |||:::
Db     107426 ------CTTGAATTGGCCTATTTGCACTCGTTGGCGGTTGATTATCCCAAAACGGGAGTG 107373
```

Search completed: March 31, 2004, 11:00:02
Job time : 3778.57 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:         March 31, 2004, 02:00:42 ; Search time 340.609 Seconds
                                           (without alignments)
                                           7146.671 Million cell updates/sec

Title:          US-09-913-878A-2_COPY_710_1282
Perfect score:  3010
Sequence:       1 GVGRMSRSVAKRIRDVLGLG..........EKWCAITPEAMDKSGANYDS 573

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:      6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09913878/runat_29032004_082007_8357/app_query.fasta_1.2254
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878_@CGN_1_1_797_@runat_29032004_082007_8357 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      N_Geneseq_29Jan04:*
                1:  geneseqn1980s:*
                2:  geneseqn1990s:*
                3:  geneseqn2000s:*
                4:  geneseqn2001as:*
                5:  geneseqn2001bs:*
                6:  geneseqn2002s:*
                7:  geneseqn2003as:*
                8:  geneseqn2003bs:*
                9:  geneseqn2003cs:*
                10:  geneseqn2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3010	100.0	8045	3	AAA65171	Aaa65171 Neurospor
2	341	11.3	615	7	ABZ54843	Abz54843 Aspergill
3	262	8.7	3731	4	AAS17845	Aas17845 Tomato RN
4	262	8.7	3731	4	AAS17837	Aas17837 Tomato RN
5	262	8.7	3731	4	AAD04370	Aad04370 Tomato RN
6	258	8.6	3591	4	AAH77704	Aah77704 Nucleotid
7	255	8.5	2816	3	AAC63741	Aac63741 Rice RNA-
8	250.5	8.3	3807	3	AAC63742	Aac63742 Soybean R
9	240.5	8.0	3737	3	AAC63739	Aac63739 Maize RNA
10	238.5	7.9	3901	3	AAC63740	Aac63740 Maize RNA
11	220.5	7.3	6863	4	AAH77703	Aah77703 Nucleotid
12	218	7.2	1281	3	AAC63738	Aac63738 Maize RNA
13	148.5	4.9	4818	4	ABL27117	Abl27117 Drosophil
14	148	4.9	282	6	ABL71773	Abl71773 Corn tass
15	132	4.4	274	6	ABL71722	Abl71722 Corn tass
16	126.5	4.2	3984	4	AAH45252	Aah45252 Human pol
17	126.5	4.2	7450	4	AAH45260	Aah45260 Human pol
18	125	4.2	110000	6	ABA90521_11	Continuation (12 o
19	124	4.1	274	7	ABX87270	Abx87270 Corn ear-
20	124	4.1	1701	5	AAS77554	Aas77554 DNA encod
21	123.5	4.1	16122	4	AAS22908	Aas22908 DNA encod
22	123.5	4.1	16341	4	AAS23002	Aas23002 DNA encod
23	120	4.0	9274	6	ABK13224	Abk13224 DNA encod
24	118.5	3.9	5641	6	ABZ32359	Abz32359 Candida a
25	117.5	3.9	11241	6	ABK82005	Abk82005 cDNA enco
26	117.5	3.9	11443	6	ABK82002	Abk82002 DNA encod
27	117.5	3.9	12923	1	AAN90338	Aan90338 Sequence
28	117.5	3.9	13957	6	ABK81959	Abk81959 cDNA enco
29	117.5	3.9	13957	6	ABT10904	Abt10904 Human bre
30	117.5	3.9	13957	6	ABN95786	Abn95786 Gene #228
31	117.5	3.9	13957	6	ABS69900	Abs69900 Human dys
32	117.5	3.9	13977	6	ABS70403	Abs70403 Human bon
33	116.5	3.9	2154	9	ADB69880	Adb69880 C. neofor
34	116.5	3.9	8114	4	ABL03787	Abl03787 Drosophil
35	116	3.9	6006	4	AAH41086	Aah41086 Human cal
36	115.5	3.8	3588	7	ACA28683	Aca28683 Prokaryot
37	115	3.8	9630	5	AAS87490	Aas87490 DNA encod
38	114.5	3.8	8370	5	AAS13652	Aas13652 cDNA sequ
39	114.5	3.8	12261	5	AAS13653	Aas13653 Genomic D
40	114	3.8	110000	6	ABA90521_00	Aba90521 Genomic s
41	113.5	3.8	11058	6	AAD37229	Aad37229 Human dys
42	113	3.8	2793	7	ACA43117	Aca43117 Prokaryot
43	113	3.8	6893	2	AAZ32182	Aaz32182 Human coa
44	113	3.8	8494	2	AAV52296	Aav52296 Streptoco
45	113	3.8	110000	7	ABS56454_06	Continuation (7 of

ALIGNMENTS

```
RESULT 1
AAA65171
ID   AAA65171 standard; DNA; 8045 BP.
XX
AC   AAA65171;
XX
DT   16-NOV-2000  (first entry)
XX
DE   Neurospora crassa qde-1 gene.
XX
KW   Gene silencing; quelling deficient; qde-1; al-1; ds.
XX
OS   Neurospora crassa.
XX
FH   Key             Location/Qualifiers
FT   CDS             2447. .6655
FT                   /*tag=  a
FT                   /product= "QDE-1"
XX
PN   WO200050581-A2.
XX
PD   31-AUG-2000.
XX
PF   16-FEB-2000; 2000WO-IT000048.
XX
PR   22-FEB-1999;   99IT-RM000117.
XX
PA   (UYRO-) UNIV ROMA LA SAPIENZA.
XX
PI   Macino G,  Cogoni C;
XX
DR   WPI; 2000-579171/54.
```

```
DR   P-PSDB; AAB13956.
XX
PT   Novel polynucleotide encoding a polypeptide which has a silencing
PT   activity and comprising a RNA-dependent RNA polymerase domain.
XX
PS   Claim 1; Page 31-43; 48pp; English.
XX
CC   The present sequence is the Neurospora crassa qde-1 gene. This gene has
CC   silencing activity. The qde-1 gene was isolated by mutational analysis of
CC   an al-1 transgenic strain. This strain had an albino phenotype resulting
CC   from post-transcriptional silencing of the endogenous al-1 gene.
CC   Reversion of this phenotype indicated a mutation in a silencing gene. The
CC   silencing gene, qde-1, could then be isolated. Modulation of qde-1
CC   expression may be used to inactivate genes and to silence suppression of
CC   genes
XX
SQ   Sequence 8045 BP; 2089 A; 2180 C; 1950 G; 1826 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              8.23e-283     Length:        8045
Score:                  3010.00       Matches:       573
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     3             Gaps:          0

US-09-913-878A-2_COPY_710_1282 (1-573) x AAA65171 (1-8045)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4574 GGTGTAGGCCGCATGTCGCGAAGCGTGGCCAAGAGGATACGCGATGTTCTCGGTTTGGGT 4633

Qy         21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4634 GATGTGCCCTCTGCTGTGCAAGGGCGGTTTGGTTCGGCCAAGGGAATGTGGGTTATCGAC 4693

Qy         41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4694 GTTGACGACACAGGCGATGAGGATTGGATCGAGACATACCCGTCCCAGCGCAAGTGGGAA 4753

Qy         61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4754 TGCGACTTCGTTGATAAACATCAACGTACCCTCGAAGTCCGGAGCGTGGCTTCTGAACTG 4813

Qy         81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4814 AAGTCAGCTGGTCTCAACCTACAGCTGTTACCTGTCCTGGAAGATAGAGCCAGGGACAAG 4873

Qy        101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4874 GTGAAGATGCGCCAGGCAATCGGTGACCGTCTTATCAACGATTTGCAACGACAGTTCAGC 4933

Qy        121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4934 GAGCAAAAGCATGCTTTGAATCGCCCAGTGGAATTTCGCCAATGGGTTTACGAGAGTTAT 4993

Qy        141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4994 TCCAGTCGCGCAACTCGAGTCAGCCACGGCCGTGTGCCTTTTCTTGCTGGGCTACCTGAC 5053

Qy        161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5054 AGTCAAGAGGAGACACTGAACTTCTTGATGAACAGTGGGTTCGATCCCAAGAAGCAAAAG 5113

Qy        181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5114 TACTTGCAAGACATCGCCTGGGATCTTCAAAAGCGGAAATGTGACACGTTGAAGTCCAAG 5173

Qy        201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5174 CTGAACATCCGTGTCGGTCGATCAGCATACATTTACATGATTGCCGATTTCTGGGGTGTG 5233

Qy        221 LeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGluSer 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5234 CTTGAGGAAAATGAGGTTCATGTCGGATTCTCCTCAAAGTTCAGAGACGAGGAGGAGTCT 5293

Qy        241 PheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSer 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5294 TTTACACTCCTATCGGACTGTGATGTCCTCGTGGCGCGATCCCCAGCCCATTTCCCTAGT 5353

Qy        261 AspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIle 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5354 GATATCCAACGGGTTCGAGCAGTCTTCAAGCCAGAGCTCCACAGTCTCAAGGATGTAATC 5413

Qy        281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5414 ATCTTCTCTACTAAAGGAGATGTACCGCTTGCTAAGAAGCTATCTGGTGGAGACTACGAC 5473

Qy        301 GlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGlu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5474 GGCGATATGGCCTGGGTCTGCTGGGATCCGGAGATCGTCGATGGTTTCGTCAATGCGGAA 5533

Qy        321 MetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGln 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5534 ATGCCTCTGGAGCCCGACCTGTCTAGGTACCTAAAGAAGGACAAAACGACTTTCAAACAA 5593

Qy        341 LeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIle 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5594 CTTATGGCCTCACACGGCACGGGCTCAGCGGCCAAAGAGCAGACTACATACGATATGATC 5653

Qy        361 GlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5654 CAGAAGAGCTTCCATTTCGCCCTGCAGCCCAACTTCTTGGGCATGTGCACTAACTACAAA 5713

Qy        381 GluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSer 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5714 GAAAGGCTCTGTTACATCAACAATAGTGTGTCTAACAAGCCGGCCATCATTCTTAGTTCA 5773

Qy        401 LeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrp 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5774 CTGGTGGGAAACCTCGTCGATCAGAGCAAGCAAGGTATTGTCTTTAACGAAGCAAGCTGG 5833

Qy        421 AlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLys 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5834 GCTCAATTGCGTAGGGAACTGCTTGGCGGTGCATTGTCCCTTCCTGACCCAATGTACAAG 5893

Qy        441 SerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSer 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5894 AGCGACAGTTGGCTCGGGCGCGGAGAGCCTACCCACATTATTGACTACCTGAAATTCTCC 5953

Qy        461 IleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAla 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5954 ATCGCCAGGCCTGCGATTGACAAGGAACTGGAAGCCTTCCACAATGCCATGAAAGCGGCC 6013

Qy        481 LysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhe 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6014 AAGGATACAGAAGACGGCGCTCACTTTTGGGATCCGGATCTCGCTTCCTACTACACGTTC 6073

Qy        501 PheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsn 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6074 TTCAAGGAGATTAGCGACAAGTCGCGATCGTCCGCACTGCTATTCACGACTCTGAAGAAC 6133

Qy        521 ArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAsp 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6134 CGTATCGGCGAAGTCGAGAAAGAATATGGCAGGTTGGTCAAAAACAAGGAGATGAGAGAC 6193

Qy        541 SerLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThr 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6194 AGCAAGGACCCCTACCCTGTCCGCGTCAACCAGGTTTATGAAAAATGGTGCGCCATCACG 6253

Qy        561 ProGluAlaMetAspLysSerGlyAlaAsnTyrAspSer 573
              |||||||||||||||||||||||||||||||||||||||
Db       6254 CCTGAGGCGATGGACAAATCCGGAGCAAATTATGATTCT 6292

RESULT 2
ABZ54843
```

```
ID   ABZ54843 standard; cDNA; 615 BP.
XX
AC   ABZ54843;
XX
DT   28-MAR-2003  (first entry)
XX
DE   Aspergillus oryzae polynucleotide SEQ ID NO 3956.
XX
KW   Aspergillus oryzae; fermentation; fungus; industrial; EST;
KW   expressed sequence tag; gene; ss.
XX
OS   Aspergillus oryzae.
XX
PN   WO200279476-A1.
XX
PD   10-OCT-2002.
XX
PF   22-MAR-2002; 2002WO-IB000890.
XX
PR   30-MAR-2001; 2001JP-00098371.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (NARE-) NAT RES INST BREWING.
PA   (NORQ ) NAT FOOD RES INST MIN AGRIC.
XX
PI   Machida M,  Akita O,  Kashiwagi Y,  Kitamoto K,  Horiuchi H;
PI   Takeuchi M,  Kobayashi T,  Kitamoto N,  Gomi K,  Abe K;
XX
DR   WPI; 2003-046817/04.
XX
PT   Detection of expression of specific Aspergillus genes for monitoring the
PT   fermentation and growth conditions of the fungus, using DNA probes.
XX
PS   Claim 1; SEQ ID NO 3956; 48pp + Sequence Listing; Japanese.
XX
CC   The invention relates to a polynucleotide having any of 6006 specific
CC   sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
CC   specific culture conditions including one or more of eutrophic,
CC   oligotrophic, solid, early germination, alkaline, high temperature, low
CC   temperature or maltose culture or polynucleotides stringently hybridising
CC   to these sequences. The polynucleotides are useful for monitoring the
CC   progress of fermentation and the growth conditions of a fungus,
CC   especially of Aspergillus oryzae which is widely used in industrial
CC   fermentation. Also monitoring for fungal contamination. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 615 BP; 151 A; 148 C; 162 G; 154 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              8.49e-24      Length:          615
Score:                  341.00        Matches:         89
Percent Similarity:     54.75%        Conservative:    32
Best Local Similarity:  40.27%        Mismatches:      80
Query Match:            11.33%        Indels:          21
DB:                     7             Gaps:            6

US-09-913-878A-2_COPY_710_1282 (1-573) x ABZ54843 (1-615)

Qy        226 ValHisValGlyPheSerSerLysPheArg-AspGluGluGluSerPhe------ThrLe 243
              ||||||   |||||||||||   :::    |||  |||     |||          ||
Db          1 GTTCACTTTGGCTTCTCCAGTAACTGGCTGTGATCCTGAAGGCCAATTCGAAGACAATCT 60

Qy        243 uLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGl 263
              ||||         ||||||||||||||||   |||||||||  |||||||||||||
Db         61 ACTCGACGGTATGGACGTTCTTGTAGCTCGGCTCCCCGCACATGTGCCGTCTGATATCCA 120

Qy        263 nArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSe 283
              ||||  |||||||||:::||||||||||||         ||||||||||||:::|||
Db        121 GCGACGAAGAGCTGTGTGGAAGCCAGAGCTCCGGCATTTCAAAGATGTTATCGTTTTCCC 180

Qy        283 rThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMe 303
              |||   |||       ||||||||     ||||||||||||||||||||||||||
Db        181 TACTGTTGGGACAACGCCACTGGCGCACATGTTATCCGGTGGTGATTACGACGGTGATAC 240

Qy        303 tAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet---Pr 322
                 |||||||||||||        ||||||      |||     |||:::::::::
Db        241 AACGTGGGTTTGCTGGGACCAAAATATTGTCCAGAAGTTTCGTAATTCCGACCTATCCAC 300

Qy        322 oLeuGluProAspLeuSerArgTyr---LeuLysLysAspLysThrThrPheLysGlnLe 341
               :::|||                  :::   |||::::::            |||   ::
Db        301 CATGGAATACCCTGCGGAGCACTTTGGTCTTGAGGAACATAACGTGCCAATGAAAGATAT 360

Qy        341 uMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGl 361
              :                                       :::    :::   :::||
Db        361 T-----------------------------------GACTCATGGGACGAGTTTCTTCA 384

Qy        361 nLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGl 381
              |   :::|||   |||    |||        :::  ||||||  ||||||           ||
Db        385 AAGCACTTTTACGTTCAATTTAACCATGTCTAATCTGGGAAGGTGTACTGTCGAGCACGA 444

Qy        381 uArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLe 401
              |::::::   |||    :::   |||:::   :::          |||      |||:::    ||
Db        445 GAAGATATCCTAC---GACGAATCGATCGACTCGCCGAATGCCAAAGAGCTCGCCTGTCT 501

Qy        401 uValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAl 421
              |:::    :::|||||||||        |||   |||:::      ::::::   :::|||
Db        502 CCTCAGTCATTTGGTGGACGGTCGTAAAGGAGGGGTCCGCTTATCAGACAG-GCGTGGCG 560

Qy        421 aGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSe 441
                 |||:::   :::            |||         :::      |||   ||||||::
Db        561 AGCATATCGCAAAAGGGTTAGCCCTAGGGCACGGGATGTG------CCTGCGTACAAGAA 614

Qy        441 r 441
              :
Db        615 C 615

RESULT 3
AAS17845
ID   AAS17845 standard; cDNA; 3731 BP.
XX
AC   AAS17845;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Tomato RNA-directed RNA polymerase (RdRP) cDNA from clone HF.
XX
KW   Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW   cytostatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW   transgenic plant; transgenic animal; cancer; viral infection;
KW   immunoprecipitation; immunolocalisation; RdRP/HF; mutant; gene therapy;
KW   ss.
XX
OS   Lycopersicon esculentum.
XX
FH   Key             Location/Qualifiers
FT   mutation        replace(31,T)
FT                   /*tag=  b
FT   mutation        replace(32,G)
FT                   /*tag=  c
FT   CDS             194. .3535
FT                   /*tag=  a
FT                   /product= "RdRP protein"
FT                   /EC_number= "2.7.7.48"
FT   mutation        replace(2017,G)
FT                   /*tag=  d
FT   mutation        replace(2264,T)
FT                   /*tag=  e
FT   mutation        replace(3597,A)
FT                   /*tag=  f
FT   mutation        replace(3599,A)
FT                   /*tag=  g
XX
PN   US2001023067-A1.
XX
```

```
PD   20-SEP-2001.
XX
PF   08-FEB-2001; 2001US-00782874.
XX
PR   05-MAR-1997;   97US-00811583.
XX
PA   (WASS/) WASSENEGGER M.
PA   (RIED/) RIEDEL L.
XX
PI   Wassenegger M,  Riedel L,  Schiebel W,  Sanger HL;
XX
DR   WPI; 2001-595798/67.
XX
PT   New nucleic acid molecule encoding a polypeptide having the enzymatic
PT   activity of RNA-directed RNA polymerase, for modulating gene expression
PT   and treating cancer and virus infection in human and animals.
XX
PS   Example 3; Page; 34pp; English.
XX
CC   This sequence represents a cDNA encoding the tomato RNA-directed RNA
CC   polymerase (RdRP/HF) of the invention. The invention comprises the
CC   nucleic acid and protein sequences of RdRP. The protein of the invention
CC   can catalyse in vitro transcription of short single stranded RNAs into
CC   DNA molecules, this transcription can be either primed by RNA or DNA
CC   oligonucleotides or be unprimed. The protein may have cytostatic or
CC   virucide activities. The sequences of the invention may be used in gene
CC   therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA
CC   sequence and a template nucleic acid molecule derived from a nucleic acid
CC   molecule which causes a disease are useful for treating a disease caused
CC   by the undesired expression or overexpression of a nucleic acid molecule
CC   in a human, rat or mouse, by administering the molecules. This system can
CC   be used in the preparation of a pharmaceutical composition and for
CC   inhibiting expression of any desired gene by transferring the RdRP system
CC   to organisms that either lack a comparable mechanism or do not
CC   sufficiently express their own RdRP. An antibody or an antagonist or
CC   inhibitor to the protein are useful for inhibiting RNA directed RNA
CC   synthesis and for ensuring stable heterologous, gene expression in
CC   transgenic organisms. The sequence is useful for probes and/or for the
CC   control of gene expression, as primers for amplification of nucleic acid
CC   molecules and as tools for the detection of expression of the cDNA
CC   molecules. Additionally, nucleotide and protein sequences are useful for
CC   suppression of undesired gene expression in humans and animals. The RdRP
CC   is useful as a therapeutic agent for the control of cancer and virus
CC   infection in humans and animals and the antibody is useful for
CC   immunoprecipitation or immunolocalisation of the protein, identification
CC   of polypeptides interacting with it and screening expression libraries.
CC   Note: This sequence is not shown in the specification but was created by
CC   the indexer from the wild type RdRP sequence shown in AAS17837 and the
CC   information shown in table 1
XX
SQ   Sequence 3731 BP; 1062 A; 672 C; 848 G; 1149 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              5.02e-15       Length:         3731
Score:                  262.00         Matches:        127
Percent Similarity:     37.66%         Conservative:   85
Best Local Similarity:  22.56%         Mismatches:     222
Query Match:            8.70%          Indels:         130
DB:                     4              Gaps:           17

US-09-913-878A-2_COPY_710_1282 (1-573) x AAS17845 (1-3731)

Qy        1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
Db     1772 GGAATTGGTAAAATATCTGGTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCTTCAA 1831

Qy       21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
Db     1832 TATACCCCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGGTGTTGAT 1891

Qy       41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
Db     1892 CCGGATTCATCAATGAAG---------TTGTCTTTGAGAAAGAGCATGTCGAAATATGAA 1942

Qy       61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
Db     1943 TCAGACAACATA--------------AAGTTAGATGTCCTTGGA---TGGAGCAAATAT 1984

Qy       81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
Db     1985 CAGCCTTGTTATCTTAATCGTCAACTGATTACTCTCTTGTCT----------------- 2026

Qy      101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
Db     2027 ---------ACACTTGGAGTGAAAGATGAAGTTCTCGAACAGAAGCAAAAGGAAGCTGTA 2077

Qy      121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
Db     2078 GATCAGCTTGATGCTATCTTGCATGATTCTTTGAAGGCACAGGAGGCTTTGGAATTGATG 2137

Qy      141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
Db     2138 TCT-----------------------------------------------------CCTGGA 2146

Qy      161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
Db     2147 GAGAACACTAATATTCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCC 2206

Qy      181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
Db     2207 TTTCTTTCAATGATGTTGCAAACCTTCCGCGCATCCAAGTTGCTCGACTTGCGGACTAGA 2266

Qy      201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
Db     2267 TCAAGAATATTTATTCCAAATGGAAGAACAATGATGGGATGTTTGGATGAATCCAGAACC 2326

Qy      221 LeuGluGluAsnGluValHisValGlyPheSer-----------SerLysPheArgAsp 236
Db     2327 TTGGAATATGGTCAGGTGTTTGTTCAGTTTACTGGTGCTGGACATGGAGAGTTTTCTGAC 2386

Qy      237 GluGluGluSerPhe----------------------ThrLeuLeuSerAspCysAsp 248
Db     2387 GATTTACATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAAT 2446

Qy      249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
Db     2447 GTGGTTGTTGCAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTTAAAGGCTGTA 2506

Qy      269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288
Db     2507 AATGTTCGAGCGCTGCACCACATGGTAGATTGTGTTGTATTCCCTCAGAAAGGAAAAAGA 2566

Qy      289 ProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrp 308
Db     2567 CCTCATCCGAATGAATGTTCTGGGAGTGATTTGGATGGGGATATCTACTTTGTTTGCTGG 2626

Qy      309 AspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSer 328
Db     2627 GATCAAGACATGATCCCGCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCCAGC 2686

Qy      329 ArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGly 348
Db     2687 ATACAGTTGGACCATGAT------------------------------------------ 2704

Qy      349 SerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeu 368
Db     2705 ------------------GTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATT 2746

Qy      369 GlnProAsnPheLeuGlyMetCysThrAsn-----------TyrLysGluArgLeuCys 384
Db     2747 GTGAATGACAGTTTGGGAATCATAGCAAATGCCCATGTCGTATTTGCAGACAGA------ 2800

Qy      385 TyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
Db     2801 ---GAACCTGATATGGCCATGAGTGATCCATGCAAAAAACTTGCTGAGCTCTTTTCAATT 2857
```

```
Qy        405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
                 |||||      |||  |||:::                |||    |||||||||
Db       2858 GCAGTGGACTTTCCAAAGACTGGTGTTCCCGCTGAAATACCATCT------CAGTTGCGC 2911

Qy        425 ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTr 444
                                                 |||        ||| |||||
Db       2912 -----------------------------------CCTAAAGAATACCCAGACTTCATG 2935

Qy        444 pLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgPr 464
              |:::   |||                                          |||||
Db       2936 GATAAGCCGG----------------------------------------ACAAGACC 2953

Qy        464 oAlaIleAspLysGlu------LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAs 482
               ||||||  ::::::       ||||||:::|||         :::|||
Db       2954 AGCTATATCTCAGAAAGAGTTATTGGAAAGCTTTTC-AGGAAAGTGAAG----------- 3001

Qy        482 pThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLy 502
                     |||   |||            |||     :::|||     ||||||    ::
Db       3002 -------GACAAAGCA-----------CCTCAGGCTAGCTCTATCGCGACCTTCACAAG 3042

Qy        502 sGluIleSerAspLysSerArgSerSerAlaLeu------------------------- 513
              :::::::::  :::|||      :::    :::
Db       3043 AGATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACAT 3102

Qy        514 ---------------LeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGl 528
                                |||       |||:::::|||    :::
Db       3103 TGACGAAGCTTTTGACTACAAAACTGAATATGACAACAAGCTGGGTAATTTAATGGACTA 3162

Qy        528 uTyrGly 530
               ||||||
Db       3163 CTATGGC 3169
```

RESULT 4
AAS17837

```
ID   AAS17837 standard; cDNA; 3731 BP.
XX
AC   AAS17837;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Tomato RNA-directed RNA polymerase (RdRP) cDNA.
XX
KW   Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW   cytostatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW   transgenic plant; transgenic animal; cancer; viral infection;
KW   immunoprecipitation; immunolocalisation; ss; gene therapy.
XX
OS   Lycopersicon esculentum.
XX
FH   Key             Location/Qualifiers
FT   CDS             194..3538
FT                   /*tag=  a
FT                   /product= "RdRP protein"
FT                   /EC_number= "2.7.7.48"
XX
PN   US2001023067-A1.
XX
PD   20-SEP-2001.
XX
PF   08-FEB-2001; 2001US-00782874.
XX
PR   05-MAR-1997;   97US-00811583.
XX
PA   (WASS/) WASSENEGGER M.
PA   (RIED/) RIEDEL L.
XX
PI   Wassenegger M,  Riedel L,  Schiebel W,  Sanger HL;
XX
DR   WPI; 2001-595798/67.
DR   P-PSDB; AAU10006.
XX
PT   New nucleic acid molecule encoding a polypeptide having the enzymatic
PT   activity of RNA-directed RNA polymerase, for modulating gene expression
PT   and treating cancer and virus infection in human and animals.
XX
PS   Claim 1; Page 15-19; 34pp; English.
XX
CC   This sequence represents a cDNA encoding the tomato RNA-directed RNA
CC   polymerase (RdRP) protein of the invention. The invention comprises the
CC   nucleic acid and protein sequences of RdRP. The protein of the invention
CC   can catalyse in vitro transcription of short single stranded RNAs into
CC   DNA molecules, this transcription can be either primed by RNA or DNA
CC   oligonucleotides or be unprimed. The protein may have cytostatic or
CC   virucide activities. The sequences of the invention may be used in gene
CC   therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA
CC   sequence and a template nucleic acid molecule derived from a nucleic acid
CC   molecule which causes a disease are useful for treating a disease caused
CC   by the undesired expression or overexpression of a nucleic acid molecule
CC   in a human, rat or mouse, by administering the molecules. This system can
CC   be used in the preparation of a pharmaceutical composition and for
CC   inhibiting expression of any desired gene by transferring the RdRP system
CC   to organisms that either lack a comparable mechanism or do not
CC   sufficiently express their own RdRP. An antibody or an antagonist or
CC   inhibitor to the protein are useful for inhibiting RNA directed RNA
CC   synthesis and for ensuring stable heterologous, gene expression in
CC   transgenic organisms. The sequence is useful for probes and/or for the
CC   control of gene expression, as primers for amplification of nucleic acid
CC   molecules and as tools for the detection of expression of the cDNA
CC   molecules. Additionally, nucleotide and protein sequences are useful for
CC   suppression of undesired gene expression in humans and animals. The RdRP
CC   is useful as a therapeutic agent for the control of cancer and virus
CC   infection in humans and animals and the antibody is useful for
CC   immunoprecipitation or immunolocalisation of the protein, identification
CC   of polypeptides interacting with it and screening expression libraries
XX
SQ   Sequence 3731 BP; 1064 A; 669 C; 849 G; 1149 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:              5.02e-15       Length:         3731
Score:                  262.00         Matches:        127
Percent Similarity:     37.66%         Conservative:   85
Best Local Similarity:  22.56%         Mismatches:     222
Query Match:            8.70%          Indels:         130
DB:                     4              Gaps:           17

US-09-913-878A-2_COPY_710_1282 (1-573) x AAS17837 (1-3731)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
              |||:::|||::::::|||                |||   |||:::          ||||||
Db       1772 GGAATTGGTAAAATATCTGGTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCTTCAA 1831

Qy         21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
                 ||||||||    |||   |||:::|||        ||||||:::         :::|||
Db       1832 TATACCCCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGGTGTTGAT 1891

Qy         41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
                 |||   :::                  :::         |||      |||:::|||
Db       1892 CCGGATTCATCAATGAAG---------TTGTCTTTGAGAAAGAGCATGTCGAAATATGAA 1942

Qy         61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
                 |||   :::                  |||:::|||            |||:::
Db       1943 TCAGACAACATA---------------AAGTTAGATGTCCTTGGA---TGGAGCAAATAT 1984

Qy         81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
              :::         ||||||   ||||||:::   :::|||
Db       1985 CAGCCTTGTTATCTTAATCGTCAACTGATTACGCTCTTGTCT------------------ 2026

Qy        101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
                          :::   |||   ::::::            |||::::::
Db       2027 ---------ACACTTGGAGTGAAAGATGAAGTTCTCGAACAGAAGCAAAAGGAAGCTGTA 2077

Qy        121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
              :::|||      |||:::                     |||         |||
Db       2078 GATCAGCTTGATGCTATCTTGCATGATTCTTTGAAGGCACAGGAGGCTTTGGAATTGATG 2137

Qy        141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
```

```
                |||                                                  |||
Db       2138 TCT----------------------------------------------------CCTGGA 2146

Qy        161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
                              |||        ::::::|||   |.|:::   |||        :::
Db       2147 GAGAACACTAATATTCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCC 2206

Qy        181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
              :::|||      :::           :::       |||        |||:::::::::
Db       2207 TTTCTTTCAATGATGTTGCAAACCTTCCGCGCATCCAAGTTGCTCGATTTGCGGACTAGA 2266

Qy        201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
                    |||   :::           :::             |||
Db       2267 TCAAGAATATTTATTCCAAATGGAAGAACAATGATGGGATGTTTGGATGAATCCAGAACC 2326

Qy        221 LeuGluGluAsnGluValHisValGlyPheSer-----------SerLysPheArgAsp 236
              ||||||      :::|||   |||   |||:::              :::|||   |||
Db       2327 TTGGAATATGGTCAGGTGTTTGTTCAGTTTACTGGTGCTGGACATGGAGAGTTTTCTGAC 2386

Qy        237 GluGluGluSerPhe-----------------------ThrLeuLeuSerAspCysAsp 248
              :::         |||                           :::         :::
Db       2387 GATTTACATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAAT 2446

Qy        249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
              |||:::||||||:::::|||            |||   ||||||:::   ::::::||||||
Db       2447 GTGGTTGTTGCAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTTAAAGGCTGTA 2506

Qy        269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288
                       ||||||   :::   |||   ::::::|||         ||||||
Db       2507 AATGTTCGAGCGCTGCACCACATGGTAGATTGTGTTGTATTCCCTCAGAAAGGAAAAAGA 2566

Qy        289 ProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrp 308
              |||         :::   ||||||   |||   |||||||||:::   :::||||||||
Db       2567 CCTCATCCGAATGAATGTTCTGGGAGTGATTTGGATGGGGATATCTACTTTGTTTGCTGG 2626

Qy        309 AspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSer 328
              |||   :::::::::                      |||         |||      |||
Db       2627 GATCAAGACATGATCCCGCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCCAGC 2686

Qy        329 ArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGly 348
                    |||         |.|
Db       2687 ATACAGTTGGACCATGAT------------------------------------------ 2704

Qy        349 SerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeu 368
                                        |||   :::   :::::::::   |||         :::
Db       2705 ------------------GTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATT 2746

Qy        369 GlnProAsnPheLeuGlyMetCysThrAsn------------TyrLysGluArgLeuCys 384
                    :::   ||||||:::            |||               :::   :::|||
Db       2747 GTGAATGACAGTTTGGGAATCATAGCAAATGCCCATGTCGTATTTGCAGACAGA------ 2800

Qy        385 TyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
                       :::         :::   |||         |||:::   |||
Db       2801 ---GAACCTGATATGGCCATGAGTGATCCATGCAAAAAACTTGCTGAGCTCTTTTCAATT 2857

Qy        405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
                 ||||||      |||   |||:::                |||      ||||||||
Db       2858 GCAGTGGACTTTCCAAAGACTGGTGTTCCCGCTGAAATACCATCT------CAGTTGCGC 2911

Qy        425 ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTr 444
                                                        |||         ||| |||||
Db       2912 ------------------------------------CCTAAAGAATACCCAGACTTCATG 2935

Qy        444 pLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgPr 464
              |:::   |||                                                |||||
Db       2936 GATAAGCCGG----------------------------------------ACAAGACC 2953

Qy        464 oAlaIleAspLysGlu------LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAs 482
              ||||||     ::::::      ||||||:::|||            :::|||
Db       2954 AGCTATATCTCAGAAAGAGTTATTGGAAAGCTTTTC-AGGAAAGTGAAG----------- 3001

Qy        482 pThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLy 502
                       |||   |||                  |||   :::|||      ||||||  ::

Db       3002 -------GACAAAGCA-----------CCTCAGGCTAGCTCTATCGCGACCTTCACAAG 3042

Qy        502 sGluIleSerAspLysSerArgSerSerAlaLeu------------------------- 513
               :::::::::   :::|||      :::    :::
Db       3043 AGATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACAT 3102

Qy        514 ----------------LeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGl 528
                                |||            |||::::::|||      :::
Db       3103 TGACGAAGCTTTTGACTACAAAACTGAATATGACAACAAGCTGGGTAATTTAATGGACTA 3162

Qy        528 uTyrGly 530
                ||||||
Db       3163 CTATGGC 3169
```

RESULT 5

```
AAD04370
ID   AAD04370 standard; cDNA; 3731 BP.
XX
AC   AAD04370;
XX
DT   04-JUL-2001  (first entry)
XX
DE   Tomato RNA-directed RNA polymerase (RdRP) cDNA.
XX
KW   Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW   transgenic plant; tissue culture; plant breeding; therapy; C-protein; ss.
XX
OS   Lycopersicon esculentum.
XX
FH   Key             Location/Qualifiers
FT   CDS             194. .3538
FT                   /*tag=  a
FT                   /product= "Tomato C-protein having RNA-directed RNA
FT                   polymerase (RdRP) activity"
XX
PN   US6218142-B1.
XX
PD   17-APR-2001.
XX
PF   05-MAR-1997;   97US-00811583.
XX
PR   05-MAR-1997;   97US-00811583.
XX
PA   (WASS/) WASSENEGGER M.
PA   (RIED/) RIEDEL L.
XX
PI   Wassenegger M,  Riedel L,  Schiebel W,  Sanger HL;
XX
DR   WPI; 2001-289830/30.
DR   P-PSDB; AAE00897.
XX
PT   New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT   polymerase enzymatic activity, useful in modulating gene expression in
PT   plants, humans and animals, as well as in plant cell/tissue cultures or
PT   plant breeding.
XX
PS   Claim 1; Col 25-34; 31pp; English.
XX
CC   The present sequence is a cDNA encoding tomato C-protein having RNA-
CC   directed RNA polymerase (RdRP) activity. This protein is capable of RNA-
CC   directed RNA synthesis, thus using RNA as a template for synthesising
CC   complementary RNA molecules. RdRP nucleic acid is useful for modulating
CC   gene expression in plants, humans and animals. This may lead to various
CC   physiological, developmental and/or morphological changes. Transgenic
CC   plants containing RdRP nucleic acid is especially useful in plant cell or
CC   tissue cultures and in plant breeding. RdRP is useful in gene therapy,
CC   particularly for treating a disease that is caused by the undesirable
CC   expression or overexpression of a gene
XX
SQ   Sequence 3731 BP; 1064 A; 669 C; 849 G; 1149 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:          5.02e-15      Length:       3731
```

```
Score:                      262.00        Matches:        127
Percent Similarity:         37.66%        Conservative:   85
Best Local Similarity:      22.56%        Mismatches:     222
Query Match:                8.70%         Indels:         130
DB:                         4             Gaps:           17

US-09-913-878A-2_COPY_710_1282 (1-573) x AAD04370 (1-3731)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
                |||:::|||::::::|||          |||   |||:::          ||||||
Db       1772 GGAATTGGTAAAATATCTGGTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCTTCAA 1831

Qy         21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
                       |||||||||   |||   |||:::|||      ||||||:::      :::|||
Db       1832 TATACCCCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGGTGTTGAT 1891

Qy         41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
                 |||      :::                 :::         |||      |||:::|||
Db       1892 CCGGATTCATCAATGAAG---------TTGTCTTTGAGAAAGAGCATGTCGAAATATGAA 1942

Qy         61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
                 |||   :::                     |||:::|||         |||:::
Db       1943 TCAGACAACATA---------------AAGTTAGATGTCCTTGGA---TGGAGCAAATAT 1984

Qy         81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
              :::               ||||||   ||||||:::   :::|||
Db       1985 CAGCCTTGTTATCTTAATCGTCAACTGATTACGCTCTTGTCT------------------ 2026

Qy        101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
                                 :::   |||   ::::::               |||::::::
Db       2027 ---------ACACTTGGAGTGAAAGATGAAGTTCTCGAACAGAAGCAAAAGGAAGCTGTA 2077

Qy        121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
              :::|||      |||:::                     |||         |||
Db       2078 GATCAGCTTGATGCTATCTTGCATGATTCTTTGAAGGCACAGGAGGCTTTGGAATTGATG 2137

Qy        141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
              |||                                                  |||
Db       2138 TCT---------------------------------------------------CCTGGA 2146

Qy        161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
                       |||      ::::::|||   |||:::      |||         :::
Db       2147 GAGAACACTAATATTCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCC 2206

Qy        181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
              :::|||      :::               :::      |||      |||:::::::::
Db       2207 TTTCTTTCAATGATGTTGCAAACCTTCCGCGCATCCAAGTTGCTCGATTTGCGGACTAGA 2266

Qy        201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
                    |||   :::                     :::         |||
Db       2267 TCAAGAATATTTATTCCAAATGGAAGAACAATGATGGGATGTTTGGATGAATCCAGAACC 2326

Qy        221 LeuGluGluAsnGluValHisValGlyPheSer-----------SerLysPheArgAsp 236
              ||||||      :::|||   |||   |||:::                  :::|||   |||
Db       2327 TTGGAATATGGTCAGGTGTTTGTTCAGTTTACTGGTGCTGGACATGGAGAGTTTTCTGAC 2386

Qy        237 GluGluGluSerPhe-----------------------ThrLeuLeuSerAspCysAsp 248
              :::         |||                               :::         :::
Db       2387 GATTTACATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAAT 2446

Qy        249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
              |||:::||||||:::::::|||               |||   ||||||:::   ::::::||||||
Db       2447 GTGGTTGTTGCAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTTAAAGGCTGTA 2506

Qy        269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288
                                ||||||   :::   |||   ::::::|||      ||||||
Db       2507 AATGTTCGAGCGCTGCACCACATGGTAGATTGTGTTGTATTCCCTCAGAAAGGAAAAAGA 2566

Qy        289 ProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrp 308
              |||         :::      ||||||   |||   ||||||||||:::   :::|||||||||
Db       2567 CCTCATCCGAATGAATGTTCTGGGAGTGATTTGGATGGGGATATCTACTTTGTTTGCTGG 2626

Qy        309 AspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSer 328
              |||   ::::::::                                    |||         |||         |||
Db       2627 GATCAAGACATGATCCCGCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCCAGC 2686

Qy        329 ArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGly 348
                    |||      |||
Db       2687 ATACAGTTGGACCATGAT------------------------------------------ 2704

Qy        349 SerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeu 368
                                |||   :::   :::::::::   |||               :::
Db       2705 -----------------GTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATT 2746

Qy        369 GlnProAsnPheLeuGlyMetCysThrAsn-----------TyrLysGluArgLeuCys 384
                 :::   ||||||:::      |||               :::   :::|||
Db       2747 GTGAATGACAGTTTGGGAATCATAGCAAATGCCCATGTCGTATTTGCAGACAGA------ 2800

Qy        385 TyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
                    :::               :::   |||         |||:::      |||
Db       2801 ---GAACCTGATATGGCCATGAGTGATCCATGCAAAAAACTTGCTGAGCTCTTTTCAATT 2857

Qy        405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
                 ||||||      |||   |||:::               |||         |||||||||
Db       2858 GCAGTGGACTTTCCAAAGACTGGTGTTCCCGCTGAAATACCATCT------CAGTTGCGC 2911

Qy        425 ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTr 444
                                                    |||               |||   |||||
Db       2912 ----------------------------------CCTAAAGAATACCCAGACTTCATG 2935

Qy        444 pLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgPr 464
              |:::   |||                                               |||||
Db       2936 GATAAGCCGG------------------------------------------ACAAGACC 2953

Qy        464 oAlaIleAspLysGlu------LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAs 482
              |||||||   ::::::         ||||||:::|||         :::|||
Db       2954 AGCTATATCTCAGAAAGAGTTATTGGAAAGCTTTTC-AGGAAAGTGAAG----------- 3001

Qy        482 pThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLy 502
                    |||   |||         |||      :::|||      ||||||   ::
Db       3002 -------GACAAAGCA------------CCTCAGGCTAGCTCTATCGCGACCTTCACAAG 3042

Qy        502 sGluIleSerAspLysSerArgSerSerAlaLeu------------------------- 513
              ::::::::::   :::|||      :::   :::
Db       3043 AGATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACAT 3102

Qy        514 ----------------LeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGl 528
                                    |||            |||:::::|||         :::
Db       3103 TGACGAAGCTTTTGACTACAAAACTGAATATGACAACAAGCTGGGTAATTTAATGGACTA 3162

Qy        528 uTyrGly 530
               ||||||
Db       3163 CTATGGC 3169
```

RESULT 6
AAH77704

```
ID   AAH77704 standard; cDNA; 3591 BP.
XX
AC   AAH77704;
XX
DT   13-NOV-2001  (first entry)
XX
DE   Nucleotide sequence of a plant SGS2 polypeptide.
XX
KW   SGS2; RNA-dependent RNA polymerase; transgene silencing;
KW   transgene stability; crop plant; viral resistance; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3591
FT                   /*tag=  a
FT                   /product= "SGS2"
XX
PN   FR2804128-A1.
XX
```

```
PD   27-JUL-2001.
XX
PF   26-JAN-2000; 2000FR-00001007.
XX
PR   26-JAN-2000; 2000FR-00001007.
XX
PA   (RHOB-) RHOBIO.
PA   (INRG ) INST NAT RECH AGRONOMIQUE.
XX
PI   Beclin C,  Elmayan T,  Mourrain P,  Vaucheret H;
XX
DR   WPI; 2001-543303/61.
DR   P-PSDB; AAG67235.
XX
PT   A new plant SGS2 gene involved in encoding an RNA-dependent RNA
PT   polymerase and in transgene silencing, increases transgene stability and
PT   expression in transgenic plants when it is inactivated.
XX
PS   Claim 4; Page 32-37; 46pp; French.
XX
CC   The present sequence encodes a plant SGS2 polypeptide. SGS2 is an RNA-
CC   dependent RNA polymerase and is involved in transgene silencing.
CC   Inactivation of SGS2 is used to increase transgene stability and
CC   expression in plants, particularly crop plants, especially maize, corn,
CC   barley, sorghum, soya, sugar cane, beet, tobacco or cotton plants.
CC   Overexpression of SGS2 can be used to increase resistance to viral
CC   infection in plants
XX
SQ   Sequence 3591 BP; 1022 A; 714 C; 858 G; 997 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               1.17e-14       Length:        3591
Score:                   258.00         Matches:       140
Percent Similarity:      37.58%         Conservative:  99
Best Local Similarity:   22.01%         Mismatches:    247
Query Match:             8.57%          Indels:        150
DB:                      4              Gaps:          28

US-09-913-878A-2_COPY_710_1282 (1-573) x AAH77704 (1-3591)

Qy        1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
Db     1759 GGAATTGGTACAATCACACCTGACCTCGCTGACGAAGTAATGGAGAAACTTAAGTTG--- 1815

Qy       21 AspVal---------ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrp 37
Db     1816 GATGTGCACTACAGCCCTTGTGCTTATCAGATACGTTACGCAGGTTTCAAAGGGGTT--- 1872

Qy       38 ValIleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln--- 56
Db     1873 -----------------------------------GTTGCTCGTTGGCCATCAAAAAGT 1896

Qy       57 --------------------ArgLysTrpGluCysAspPheValAspLysHisGlnArg 69
Db     1897 GATGGAATCAGGCTAGCCCTTCGAGACAGTATGAAGAAGTTCTTTTCCAAACAT---ACG 1953

Qy       70 ThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeu 89
Db     1954 ATCTTGGAGATCTGTTCC---TGGACGAGGTTTCAACCTGGGTTCTTAAATCGGCAGATA 2010

Qy       90 LeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAsp 109
Db     2011 ATTACCCTTCTATCC--------------------------GTACTAGGTGTTCCGGAT 2043

Qy      110 ArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgPro 129
Db     2044 GAAATATTCTGGGATATGCAG---------GAATCCATGCTCTATAAACTGAACCGCATC 2094

Qy      130 ValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHis 149
Db     2095 CTTGATGATACAGATGTGGCATTTGAAGTTCTCACGGCATCATGTGCT------------ 2142

Qy      150 GlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeu 169
Db     2143 --------------------------------GAACAGGGAAACACTGCAGCTATCATG 2169

Qy      170 MetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeu 189
Db     2170 CTTAGTGCAGGTTTCAAACCAAAAACCGAGCCGCATCTACGCGGGATGTTGTCTTCAGTC 2229

Qy      190 GlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAla 209
Db     2230 AGAATTGCACAACTCTGGGGTCTCAGAGAAAAATCTCGTATTTTTGTTACTTCAGGAAGG 2289

Qy      210 TyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGly 229
Db     2290 TGGCTAATGGGTTGCCTAGACGAAGCAGGGATACTTGAACATGGCCAATGCTTTATTCAA 2349

Qy      230 PheSer------------------------------------SerLysPheArgAspGlu 237
Db     2350 GTCTCTAAACCGTCTATAGAAAATTGTTTCTCCAAACATGGGTCTCGTTTTAAGGAGACA 2409

Qy      238 GluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHis 257
Db     2410 AAGACAGATCTGGAAGTAGTTAAAGGCTAT---GTAGCCATTGCTAAGAATCCTTGTCTT 2466

Qy      258 PheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLys 277
Db     2467 CACCCAGGGGATGTAAGGATTTTAGAAGCTGTTGATGTACCCCAGCTGCATCACATGTAT 2526

Qy      278 AspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGly 297
Db     2527 GACTGCCTTATTTCCCTCAGAAAGGTGATAGGCCGCATACAAACGAAGCTTCTGGCAGT 2586

Qy      298 AspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheVal 317
Db     2587 GACCTTGACGGGGACCTGTACTTTGTGGCTTGGGATCAGAAACTCATC------------ 2634

Qy      318 AsnAlaGluMetProLeuGluProAspLeuSerArgTyr------LeuLysLysAspLys 335
Db     2635 -----------------CCTCCCAACAGGAAAAGCTATCCGGCCATGCATTATGATGCA 2676

Qy      336 ThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThr 355
Db     2677 GCTGAAGAGAAGAGTTTA-----------------GGCCGTGCTGTCAACCACCAGGAC 2718

Qy      356 ThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMet 375
Db     2719 ATAATCGATTTCTTTGCAAGAAAC------------TTGGCGAATGAGCAGTTGGGCACA 2766

Qy      376 CysThrAsn------------TyrLysGluArgLeuCysTyrIleAsnAsnSerValSer 391
Db     2767 ATTTGCAATGCACACGTCGTTCATGCTGATAGAAGTGAGTAT---------GGAGCCATG 2817

Qy      392 AsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGln 411
Db     2818 GACGAAGAATGTTTGCTACTGGCAGAACTAGCTGCCACTGCAGTTGATTTCCCAAAGACA 2877

Qy      412 GlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAla 431
Db     2878 GGGAAAATTGTGTCAATGCCC------TTCCACCTAAAACCAAAACTCTACCCAGATTTC 2931

Qy      432 LeuSerLeuProAsp---ProMetTyrLysSerAspSerTrpLeuGlyArg--------- 447
Db     2932 ATGGGAAAAGAAGACTACCAAACTTACAAGTCGAACAAAATCTTGGGTCGGCTTTACAGA 2991

Qy      448 ---------------------------------------------------GlyGluPro 450
Db     2992 CGGGTAAAAGAGGTTTATGATGAAGATGCAGAAGCTTCCTCAGAAGAAAGCACAGACCCA 3051

Qy      451 ThrHisIle---------------IleAspTyrLeuLysPheSerIleAlaArgPro--- 464
Db     3052 AGTGCCATCCCTTATGACGCTGTTCTTGAAATACCGGGATTTGAAGATTTGATCCCTGAG 3111

Qy      465 ------------------AlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLys 478
```

```
Db       3112 GCATGGGGTCACAAATGTTTGTACGACGGGCAACTCATTGGTCTTCTTGGGCAATACAAG 3171

Qy        479 AlaAlaLysAspThrGlu---AspGlyAlaHisPheTrpAspProAspLeuAlaSerTyr 497

Db       3172 GTGCAGAAAGAGGAAGAGATTGTGACGGGTCACATCTGGTCC---------ATGCCCAAA 3222

Qy        498 TyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThr 517

Db       3223 TACACAAGCAAGAAACAAGGCGAACTGAAAGAAAGA------------------------ 3258

Qy        518 LeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGlu 537

Db       3259 CTGAAGCATTCTTATAATTCCCTTAAAAAGGAGTTC---AGGAAAGTATTTGAGGAAACA 3315

Qy        538 MetArgAspSerLysAspProTyrProValArgValAsnGlnValTyrGluLys------ 555

Db       3316 ATCCCTGACCATGAAAATCTCAGCGAGGAGGAAAAAAACATCTTGTATGAGAAGAAGGCT 3375

Qy        556 ------TrpCysAlaIleThr------ProGluAlaMetAspLysSer 567

Db       3376 TCAGCTTGGTATCATGTAACTTACCATCCCGAGTGGGTGAAGAAGTCT 3423
```

RESULT 7
AAC63741
ID   AAC63741 standard; cDNA; 2816 BP.
XX
AC   AAC63741;
XX
DT   07-FEB-2001 (first entry)
XX
DE   Rice RNA-directed RNA polymerase FIS sequence, SEQ ID NO: 9.
XX
KW   Rice; plant; RNA-directed RNA polymerase; FIS; gene mapping; gene marker;
KW   plant virus resistance; plant breeding; ss.
XX
OS   Oryza sativa.
XX
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   P-PSDB; AAB28532.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 2; Page 45-46; 62pp; English.
XX
CC   The present sequence is one of a number of cDNA molecules which encode
CC   plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC   probes for genetically and physically mapping genes, and as markers for
CC   traits linked to those genes. They are useful for controlling gene
CC   expression and provide mechanisms to engineer plant virus resistance.
CC   They are also useful for plant breeding to develop lines with desired
CC   phenotypes
XX
SQ   Sequence 2816 BP; 782 A; 572 C; 725 G; 736 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:	1.66e-14	Length:	2816
Score:	255.00	Matches:	106
Percent Similarity:	39.33%	Conservative:	82
Best Local Similarity:	22.18%	Mismatches:	188
Query Match:	8.47%	Indels:	102
DB:	3	Gaps:	16

US-09-913-878A-2_COPY_710_1282 (1-573) x AAC63741 (1-2816)

```
Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20

Db        673 GGAATTGGCAAAATTACCCCAGATCTTGCATTGGAAGTTGCCGAGAGGCTGCAGCTGACA 732

Qy         21 Asp---ValProSerAlaValGlnGlyArgPheGlySerAlaLysGly---------Met 36

Db        733 GATAACCCGCCATCTGCTTATCAGATCAGGTTTGCTGGCTTCAAGGGTGTCATAGCTGTC 792

Qy         37 TrpValIleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln 56

Db        793 TGG------------CAAGGACATGGTGATGGGACACGGCTTTTCCTGAGGCCAAGCATG 840

Qy         57 ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerVal 76

Db        841 AGGAAGTTTGAGTCTAACCATTTG-----------------------GTGTTAGGGGTG 876

Qy         77 AlaSer------GluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGlu 94

Db        877 GTCTCCTGGACAAAGTTCCAGCCAGGATTCTTAAATCGACAGATTATAATATTGCTATCC 936

Qy         95 AspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAsp 114

Db        937 TCA--------------------------CTGAATGTCCCAGATTCTATCTTTTGGCAA 969

Qy        115 LeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGln 134

Db        970 ATGCAAGAGACCATGCTTTCTAAC---------CTCAACAATATTCTATCAGACAGAGAT 1020

Qy        135 TrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPhe 154

Db       1021 GTTGCTTTTGAGGTTTTAACAACTTCATGTGCT-------------------------- 1053

Qy        155 LeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPhe 174

Db       1054 ------------------GATGATGGAAACACTGCAGCATTGATGCTCAGTGCTGGCTTT 1095

Qy        175 AspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCys 194

Db       1096 GAACCTAGAACTGAACCACACTTGAAAGCAATGCTCTTGGCTATAAGGTCTGCACAATTG 1155

Qy        195 AspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIle 214

Db       1156 CAGGATCTTTTGGAAAAAGCAAGGATATTTGTGCCAAAGGGAAGGTGGTTGATGGGCTGT 1215

Qy        215 AlaAspPheTrpGlyValLeuGluGlu-------------------------------- 223

Db       1216 CTTGATGAGCTTGGGGTTCTTGAGCAAGGGCAGTGCTTTATTCGGGCAACAGTTCCATCA 1275

Qy        224 ---AsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheThr 242

Db       1276 TTGAATAGTTATTTGTTAAGCATGGGTCAAGATTTTCATCAACAGATAAAAACACAGAG 1335

Qy        243 LeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIle 262

Db       1336 GTCATTTTGGGT---ACTGTGGTAATAGCAAAGAATCCCTGTCTTCATCCAGGGGATGTC 1392

Qy        263 GlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePhe 282

Db       1393 CGCATCCTTGAAGCAGTTGATGTGCCCGAACTGCATCATCTGGTTGATTGTTTGGTGTTC 1452

Qy        283 SerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAsp 302

Db       1453 CCCCAGAAAGGTGAGAGGCCACATGCTAACGAGGCATCTGGGAGCGATCTTGATGGGGAT 1512

Qy        303 MetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetPro 322
```

```
Db       1513 CTCTACTTTGTGACATGGGATGAGAAACTTATACCTCCAGGCAAGAAGAGCTGGAACCCT 1572

Qy        323 LeuGlu---------ProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLys 339
              ::::::          |||:::                                  |||
Db       1573 ATGGACTACTCCCCACCTGAAGCA--------------------------------AAA 1599

Qy        340 GlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMet 359
              ||||||                          ::::::  :::   :::|||:::
Db       1600 CAACTT--------------------------CCGCGCCAAGTATCTCAACATGATATC 1632

Qy        360 IleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyr 379
               |||                    :::     |||       :::|||        :::
Db       1633 ATTGATTTCTTCTTAAAGAACATGATAAGTGAGAATCTTGGTAGGATCTGTAATGCTCAT 1692

Qy        380 LysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSer 399
                                                ::::::      |||   |||:::
Db       1693 GTTGTTCATGCTGATCTTAGCGAGTATGGTGCAATGGATGAGAAGTGTATTCACTTAGCT 1752

Qy        400 SerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSer 419
                 |||          ||||||      |||   |||
Db       1753 GAGCTAGCAGCAACTGCCGTGGACTTCCCCAAGACTGGC--------------------- 1791

Qy        420 TrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSer---------LeuProAsp 436
                          :::     ::::::         |||              ||||||
Db       1792 ------------AAACTTGCGATAATGCCACCACACCTTAAACCAAAAGTCTACCCTGAC 1839

Qy        437 ProMet--------------------TyrLysSerAspSerTrpLeuGlyArg 447
                 |||                  |||||||||:::     |||||||||
Db       1840 TTCATGGGAAAGGAAGATGGACAATCTTATAAATCAGAAAAGATTCTTGGAAGG 1893

RESULT 8
AAC63742
ID   AAC63742 standard; cDNA; 3807 BP.
XX
AC   AAC63742;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Soybean RNA-directed RNA polymerase FIS contig, SEQ ID NO: 11.
XX
KW   Soybean; plant; RNA-directed RNA polymerase; FIS; gene mapping;
KW   gene marker; plant virus resistance; plant breeding; ss.
XX
OS   Glycine max.
XX
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT,  Orozco EM,  Weng Z,  Sakai H,  Cahoon RE,  Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   P-PSDB; AAB28533.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 2; Page 49-50; 62pp; English.
XX
CC   The present sequence is one of a number of cDNA molecules which encode
CC   plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC   probes for genetically and physically mapping genes, and as markers for
CC   traits linked to those genes. They are useful for controlling gene
CC   expression and provide mechanisms to engineer plant virus resistance.
CC   They are also useful for plant breeding to develop lines with desired
CC   phenotypes
XX
SQ   Sequence 3807 BP; 1132 A; 662 C; 826 G; 1185 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.:              6.85e-14       Length:          3807
Score:                  250.50         Matches:         132
Percent Similarity:     38.25%         Conservative:    91
Best Local Similarity:  22.64%         Mismatches:      240
Query Match:            8.32%          Indels:          121
DB:                     3              Gaps:            23

US-09-913-878A-2_COPY_710_1282 (1-573) x AAC63742 (1-3807)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
              |||:::|||::::::|||            |||:::::::::             |||
Db       1818 GGAATTGGGAAAATATCTCTTGAATTTGCCCAGAAAGTGGCTAAAAAATGTGGTTATGAT 1877

Qy         21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
                    |||||||||   |||   |||:::|||        ||||||:::      :::|||
Db       1878 TGCACTCCATCTGCCTTTCAGATTCGATATGGTGGGTACAAAGGAGTTGTGGCTGTTGAC 1937

Qy         41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
                                      :::               |||   ||||||::::::
Db       1938 CCTAAATCATGCTACAAG---------TTATCACTGAGGAAGAGCATGCGGAAGTATGAT 1988

Qy         61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
                 |||                     |||:::|||   :::      |||:::
Db       1989 TCAGAT---------------AACACAAAGTTAGATGTTTTGGCC---CGTAGTAAGTTT 2030

Qy         81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
              :::         ||||||     ||||||:::   :::|||
Db       2031 CAGCCATGTTATCTGAATCGGCAGTTAATTTCTCTCTTATCC------------------ 2072

Qy        101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
                                |||   |||     :::               ||||||:::
Db       2073 ---------ACTCTTGGTATCAAGGATGATGTTTTGAGAAAAAACAAAGAGAAACTGTT 2123

Qy        121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
                 |||   :::   :::                        |||   |||   :::
Db       2124 AATCAACTGAACACTATACTAACAGATTCATTAAAGGCTCAGGAAGTTCTGGACTTAATG 2183

Qy        141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
              |||:::                     |||   :::
Db       2184 TCTGCT---------------------GGAGAGATC------------------------ 2198

Qy        161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
                             |||      ::::::      |||:::   |||   ::::::
Db       2199 ------ACTAATGTTCTGAAGGAGATGCTCATTTGTGGATACAAGCCTAATGAAGAACCA 2252

Qy        181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
              :::|||      :::         :::         |||         |||:::   |||
Db       2253 TTCCTTTCAATGATGCTTCAAACATTTAGGGCATCAAAACTTTTGGAATTGCGACTTAAA 2312

Qy        201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
                    |||   :::   :::         :::      |||
Db       2313 TCTAGGATCTTTATTCCAAAAGGAAGAGCAATGATGGGATGTCTAGATGAAACTAGAACC 2372

Qy        221 LeuGluGluAsnGluValHisValGlyPheSerSer-----------LysPheArgAsp 236
              ||||||      :::|||   |||   ||||||:::                     |||
Db       2373 CTAGAATATGGTCAAGTATTTGTTCAGTTCTCTAACAATAGGCTGCAGAATCTATCTGAT 2432

Qy        237 GluGluGluSerPheThrLeu-----------LeuSerAspCysAspValLeuValAla 252
              :::      |||:::   |||            :::         |||:::||||||
Db       2433 GATTTTTTTTCATATGATTTGCCAAAGAATTATATGGTTAAAGGTAAGGTAGTTGTAGCA 2492

Qy        253 ArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGlu 272
              ::::::|||         |||   |||::::::   ::::::||||||      |||:::
Db       2493 AAAAAYCCCTGCTTGCACCCAGGTGATGTGCGTGTTTTACAAGCTGTGGATGTGCCAGAT 2552

Qy        273 LeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLys 292
              |||:::   :::   |||      ::::::|||         ||||||      |||
```

```
Db     2553 TTGTACCACATGGTGGACTGTGTTGTTTTCCCTCAAAAAGGACCAAGACCTCATCCAAAT 2612

Qy      293 LysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIle 312
            :::  |||||  |||  ||||||||:::  :::||||||||||  |||:::
Db     2613 GAGTGTTCGGGAAGTGATCTGGATGGAGATATCTACTTTGTTTGTTGGGACCATGAATTG 2672

Qy      313 ValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLys 332
            :::                      |||::::::|||
Db     2673 ATT---------------CCTTCTCGCCCAATTGATCCA--------------------- 2696

Qy      333 LysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLys 352
               |||  |||        :::        |||
Db     2697 ATGGACTATACTGCTCCYGCAACTGTGGAATTGGATCAT--------------------- 2735

Qy      353 GluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPhe 372
                  :::  :::::::::  |||        :::        :::
Db     2736 ---GATGTGATGATCGAGGAGGTTGAGGAGTATTTTGCCAATTACATAGTCAATGACAGT 2792

Qy      373 LeuGlyMetCysThrAsnTyr------------------LysGluArgLeuCysTyrIle 386
            ||||||:::  |||                  ||||||  |||
Db     2793 CTGGGAATAATTGCCAATGCACACACTGTCTTTGCAGATAAAGAACATTTG--------- 2843

Qy      387 AsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVal 406
                     :::          :::  |||:::  |||              |||
Db     2844 ------AAAGCAATGTCTGATCAATGTGTTAAGCTTGCAAGGTTGTTTTCAACAGCAGTT 2897

Qy      407 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 426
            |||       |||  |||:::               |||  :::      |||
Db     2898 GACTTTCCTAAAACTGGTGTTCCA-----------------GCTGTTATACCTCCTGAA 2939

Qy      427 Leu------------LeuGlyGlyAlaLeuSerLeuProAsp---ProMetTyrLysSer 441
            |||                  :::  ||||||  |||  |||||||||
Db     2940 CTTCATGTCAAAGAATATCCTGACTTCATGGAGAAGCCTGACAAACCCACATACAAATCG 2999

Qy      442 AspSerTrpLeuGlyArg--------------------------GlyGluProThrHis 452
              :::  :::|||:::
Db     3000 CATAACGTGATAGGAAAGCTCTTTAGGGAAGTGAAAGAAATATCAACAAGTGCCGGCTCA 3059

Qy      453 IleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAla 472
            |||     :::  |||     :::||||||  :::  |||  |||:::|||
Db     3060 ATTACATCCTTCACAAAGTTG---GTTGCGAGAGACTCTTACGACCATGAAATGGAAGTT 3116

Qy      473 --------------------------------PheHisAsnAlaMetLysAlaAlaLys 481
                                            :::|||              |||
Db     3117 GATGGCTTCATGGATTATGTTGATGATGCTTTCTATCACAAAACCAATTATGACTACAAG 3176

Qy      482 AspThrGluAspGlyAlaHisPheTrpAspProAspLeuAla---------------Ser 496
                  :::                |||||  :::  :::
Db     3177 TT-GGGAAATCTGATGGACTACTATGGGATCAAAACTGAAGCTGAAATCCTCGGTGGGAA 3235

Qy      497 TyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThr 516
            ||||||        |||        |||               :::
Db     3236 TATTATGAAAATGTCAAAATCTTTCAACAAAAGGAGGGATGCAGAAGC------------ 3283

Qy      517 ThrLeuLysAsnArgIleGly---GluValGluLysGluTyrGlyArg---LeuValLys 534
                     |||:::  |||  ||||||  ||||||  |||:::  ||||||:::
Db     3284 ---------AATCAATATGGCTGTGAGGTCCCTAAGGAAAGAGGCCAGGGCCTGGTTCAA 3334

Qy      535 AsnLysGlu 537
               |||:::
Db     3335 TGAAAACAG 3343
```

RESULT 9
AAC63739

```
ID   AAC63739 standard; cDNA; 3737 BP.
XX
AC   AAC63739;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Maize RNA-directed RNA polymerase EST sequence, SEQ ID NO: 5.
XX
KW   Maize; plant; RNA-directed RNA polymerase; expressed sequence tag; EST;
KW   gene mapping; gene marker; plant virus resistance; plant breeding; ss.
XX
OS   Zea mays.
XX
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT,  Orozco EM,  Weng Z,  Sakai H,  Cahoon RE,  Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   P-PSDB; AAB28530.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 2; Page 35-36; 62pp; English.
XX
CC   The present sequence is one of a number of cDNA molecules which encode
CC   plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC   probes for genetically and physically mapping genes, and as markers for
CC   traits linked to those genes. They are useful for controlling gene
CC   expression and provide mechanisms to engineer plant virus resistance.
CC   They are also useful for plant breeding to develop lines with desired
CC   phenotypes
XX
SQ   Sequence 3737 BP; 1063 A; 760 C; 863 G; 1051 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.32e-13      Length:         3737
Score:                  240.50        Matches:        137
Percent Similarity:     36.87%        Conservative:   96
Best Local Similarity:  21.68%        Mismatches:     258
Query Match:            7.99%         Indels:         141
DB:                     3             Gaps:           23

US-09-913-878A-2_COPY_710_1282 (1-573) x AAC63739 (1-3737)

Qy        1 GlyValGlyArgMetSerArgSerValAla--------------------------Lys 11
            ||||||||||::::::|||  :::  |||                          |||
Db     1709 GGAGTTGGAAAGATCTCAGCTAATTTTGCAGTGGAGGTGGCTATGAAGTGCAAATTGAAA 1768

Qy       12 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGly 31
            |||                        ||||||      |||  |||:::|||
Db     1769 CGCTTT-----------------------GCTCCTTCTGTTTTTCAGATAAGGTATGGC 1804

Qy       32 SerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGlu 51
                  ||||||:::     :::        ||||||              :::
Db     1805 GGTTACAAAGGTGTTGTCGCTGTA---------GATACAAGATCAAATCATAAGCTTTCT 1855

Qy       52 ThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 71
                     |||     |||::::::  :::                    ||||||
Db     1856 TTGAGAAAAAGCATGTCAAAGTTCCAGTCAGAAAATATC---------------ACTCTT 1900

Qy       72 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuPro 91
            :::|||  :::     |||:::  :::              ||||||  ||||||:::
Db     1901 GATGTCCTTGCATAC---AGCAAGTACCAACCATGCTTCCTGAATCGGCAGTTGATTACT 1957

Qy       92 ValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeu 111
            :::|||                                    :::  |||  :::
Db     1958 CTTCTCTCA--------------------------ACACTTGGGGTTAGCGATAATGTC 1990

Qy      112 IleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu 131
                  |||::::::           |||           |||||||||  |||
```

```
Db     1991 TTTGAGCTAAAGCAGAAGGAAGCCTTAAGGCAG---------TTGAACAGAATGGTAACT 2041

Qy      132 PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArg 151

Db     2042 ---------------------------GAACCACAGGCTGCTCGTGAAGCAGTTGAACTT 2074

Qy      152 ValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsn 171

Db     2075 ATGCCCATG---------------GGAGAGGTAACCAATGTAGTTAAAGAATTGTTGTCA 2119

Qy      172 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 191

Db     2120 TGTGGCTACCAGCCTGATCATGAGCCATATCTTTCCATGCTGCTACAAACTTTTAGAGCA 2179

Qy      192 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 211

Db     2180 TCCAAGCTTCTAGAGTTGAAAACAAAGTCAAGGATATTCATCACACAGGGGCGAGCAATG 2239

Qy      212 TyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer 231

Db     2240 ATGGGTTGCCTGGATGAAACCTGCACACTTAAGTACGGCCAGGTATTCGTCCAAGCTTCT 2299

Qy      232 SerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 251

Db     2300 TACAGTGCAGATGACCATCGCAAG---------GTCGTTGTAACTGGAAAAGTAGTTGTC 2350

Qy      252 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 271

Db     2351 GCCAAAAATCCTTGTCTCCACCCTGGTGACATACGGGTTCTCCAGGCTGTTGATGTTCCT 2410

Qy      272 GluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAla 291

Db     2411 GCTCTGCACCACTTGTTTGACTGTGTTGTCTTTCCACAGCAGGGACCAAGGCCGCACCCT 2470

Qy      292 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 311

Db     2471 AATGAGTGTTCAGGGAGTGATCTTGATGGGGACATATATTTTGTTTCTTGGGATCCACAT 2530

Qy      312 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu 331

Db     2531 CTTATTCCAAGTCGTTTGGTGGATCCTATGGACTATACTCCAGCTTCAGCAGAAACATTA 2590

Qy      332 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 351

Db     2591 GACCATGAT--------------------------------------------------- 2599

Qy      352 LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 371

Db     2600 ---------GTCACTATTGAGGAGATACAGGAGTACTTCACAAACTACATAGTTAATGAG 2650

Qy      372 PheLeuGlyMetCysThrAsnTyr------------------LysGluArgLeuCysTyr 385

Db     2651 AGTCTTGGGATTATCGCCAATGCGCATGTGGTCTTTACAGATCAGGAACGTATG------ 2704

Qy      386 IleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeu 405

Db     2705 ---------AAAGCTGAGAGTCCACCGTGCGTTCAACTGGCCAAGCTCTTCTCTATAGCT 2755

Qy      406 ValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArg 425

Db     2756 GTCGATTTCCCAAAGACTGGAGTGCCG------------------GCTCTGATTCCACAT 2797

Qy      426 GluLeuLeuGlyGlyAlaLeuSerLeuProAspProMet--------------------- 438

Db     2798 GAGCTA------CATGTCAAGGAGTATCCTGACTTCATGGAGAAACTCGACAAAGTCACC 2851

Qy      439 TyrLysSerAspSerTrpLeuGlyArg---------------GlyGluProThrHisIle 453

Db     2852 TATGAATCAAAGGGTGTGATCGGGAAGCTCTATAGGGAAATAAAGAAGCACACACCACAC 2911

Qy      454 IleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeu------Glu 471

Db     2912 ATAAAGCACTTCACGAGGGAAGTGGCAAGGCGGTCTTATGACACCGATTTGATTGTTGAT 2971

Qy      472 AlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAsp 491

Db     2972 GGCTATGAAGATTACATTACTGAGGCTATAGAGTTCAAGGAAGAGTACGATTTCAGGCTG 3031

Qy      492 ProAspLeuAlaSerTyrTyrThrPhePheLysGlu---------IleSerAspLysSer 508

Db     3032 GGTAATCTTATGGACCACTATGGCATAAAAAGTGAAGCTGAGATAATAAGTGGATGTATT 3091

Qy      509 ArgSerSerAlaLeuLeuPheThr------------------ThrLeuLysAsnArgIle 522

Db     3092 CTAAAGATGGCAAAGAATTTCACCAAGAGTAGTGATGCTGATGCAATTAGAATGGCGGTG 3151

Qy      523 GlyGluValGluLysGlu------------TyrGlyArgLeuValLysAsnLysGluMet 538

Db     3152 AGATCTTTGAGGAAAGAAGCTAGGTCGTGGTTCAATGAGATGAGCACAGGAGAGGATGGC 3211

Qy      539 ArgAspSerLysAspPro---------------TyrProValArgValAsnGlnValTyr 553

Db     3212 CAAGATGCCATGGAGGCCAAGGCCTCTGCTTGGTACCATGTTACTTATCATCAGCAGTAC 3271

Qy      554 ------------------------------------------GluLysTrpCysAla 558

Db     3272 TGGGGCAGCTACAATGAAGGGTATGATCGGCCGCATCTTATTAGCTTCCCATGGTGCGTA 3331

Qy      559 IleThrProGluAlaMetAspLysSerGlyAlaAsn 570

Db     3332 TATGACAAGCTTGTGGCCATCAAGCAGGGGAGGAAT 3367

RESULT 10
AAC63740
ID   AAC63740 standard; cDNA; 3901 BP.
XX
AC   AAC63740;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Maize RNA-directed RNA polymerase FIS sequence, SEQ ID NO: 7.
XX
KW   Maize; plant; RNA-directed RNA polymerase; FIS; gene mapping;
KW   gene marker; plant virus resistance; plant breeding; ss.
XX
OS   Zea mays.
XX
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT,  Orozco EM,  Weng Z,  Sakai H,  Cahoon RE,  Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   P-PSDB; AAB28531.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 2; Page 40-41; 62pp; English.
XX
CC   The present sequence is one of a number of cDNA molecules which encode
CC   plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC   probes for genetically and physically mapping genes, and as markers for
CC   traits linked to those genes. They are useful for controlling gene
CC   expression and provide mechanisms to engineer plant virus resistance.
CC   They are also useful for plant breeding to develop lines with desired
```

```
CC   phenotypes
XX
SQ   Sequence 3901 BP; 928 A; 975 C; 1058 G; 935 T; 0 U; 5 Other;

Alignment Scores:
Pred. No.:              1.05e-12      Length:         3901
Score:                  238.50        Matches:        138
Percent Similarity:     36.48%        Conservative:   98
Best Local Similarity:  21.33%        Mismatches:     248
Query Match:            7.92%         Indels:         163
DB:                     3             Gaps:           26

US-09-913-878A-2_COPY_710_1282 (1-573) x AAC63740 (1-3901)

Qy        1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
Db     1857 GGAATTGGCAAGATTACGTGCGACCTTGCACTCGAAGTTGCTCAGAAGCTGCAATTGACA 1916

Qy       21 Asp---ValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIle 39
Db     1917 GATAATCCCCCATCTGCTTACCAGATTAGGTATGCAGGCTTCAAGGGT------GTTATA 1970

Qy       40 AspVal---AspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLys 58
Db     1971 TCTGTCTGGGAAGGAAAAAATGATGGGATACGACTTTCCCTGAGGCCGAGCATGCACAAG 2030

Qy       59 TrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSer 78
Db     2031 TTTGAG--------------TCTAACCATACTGTGTTAGAGGTGGTCTCG---TGGACA 2072

Qy       79 GluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArg 98
Db     2073 AAGTTTCAGCCAGGATTCTTAAATCGTCAGATTATTACATTACTGTCC----------- 2120

Qy       99 AspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGln 118
Db     2120 ------------------------------------------------------------ 2120

Qy      119 PheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGlu 138
Db     2121 -----------------TCCTTGAATGTCCCGGATGCTATCTTTGCTCAAATGCAGGAA 2162

Qy      139 SerTyrSerSerArgAlaThrArgVal---SerHisGlyArgValProPhe------Leu 155
Db     2163 GCCATGTTATCTAATCTCAACAATATTTTGTCAGACTCTGATGTTGCTTTTGACATTGTA 2222

Qy      156 AlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAsp 175
Db     2223 ACCGCCTCTTGTGCTGAGCAAGGAACCACTGCAGCACTGATGTTGAGTGCTGGCATTTCA 2282

Qy      176 ProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAsp 195
Db     2283 CCTGGAACTGAGCCACACCTGAAAGCAATGCTGTTAGCTATAAGGTCCTCACAGCTGCTA 2342

Qy      196 ThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAla 215
Db     2343 GGTCTTTTGGAGAAGACAAGGATTTTTGTGCCCAAGGGAAGGTGGTTGATGGGCTGCCTT 2402

Qy      216 AspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArg 235
Db     2403 GATGAACTTGGGATCCTTGAGCAAGGGCAGTGCTTTATCCGGGCATCATCT--------- 2453

Qy      236 AspGluGluGluSerPheThrLeuLeuSerAspCys------------------------ 247
Db     2454 -----------------CCATCACTCAATAATTGTCTGGTGAAGTATGGATCAAGATTG 2495

Qy      248 -------------------------------------AspValLeuValAlaArgSerPro 255
Db     2496 TCTGCAGCAAACACAAATGCAGAGACCATTCTGGGTACTATCGTAATGGCAAAGAATCCA 2555

Qy      256 AlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSer 275
Db     2556 TGCCTTCATCCAGGGGATGTCCGAATCCTTGAAGCTGTTGATGTGCCTGAACTGCATCAC 2615

Qy      276 LeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSer 295
Db     2616 CTTGTTGATTGCTTGGTCTTCCCCAAGAAAGGTGAGAGGCCGCACGCGAATGAAGCATCT 2675

Qy      296 GlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGly 315
Db     2676 GGGAGTGATCTTGATGGGGATCTATACTTCGTAACATGGGATGAAAACCTTATACCACCT 2735

Qy      316 PheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLys 335
Db     2736 GGTAAAAAGAGTTGGAACCCTATGGAC---------------TACTCCCCAGCTGAAGCA 2780

Qy      336 ThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThr 355
Db     2781 AAACAACTGCCACGCGCAGTATCCCAACAT----------------------------- 2810

Qy      356 ThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMet 375
Db     2811 ------GATATTGTT------GGTTTC--------------------TTCTTGAAGAAC 2837

Qy      376 CysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSer------------------ 389
Db     2838 ATGGTAAAT------GAGAAACTGGGTCCAATAAGCAATGCTCATGTTGTTCACGCTGAT 2891

Qy      390 ---------------ValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
Db     2892 ATGAGCGAGTATGGAGCAATGGATGAGAAGTGTATTCAGTTGGCAGAACTAGCAGCAACT 2951

Qy      405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
Db     2952 GCTGTGGACTTCCCCAAGACAGGCAAAATTGTGTCAATGCCA------GCATCCCTTCGA 3005

Qy      425 ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMet---TyrLysSerAspSer 443
Db     3006 CCAAAATTATATCCTGACTTCATGGGAAAGGAGGATGCTATCTCCTATAGATCAGAGAAG 3065

Qy      444 TrpLeuGlyArg------------------------------------------------ 447
Db     3066 ATCCTTGGAAGGCTTTATCGGTCAATCCAAGAAGCCTCCAGCGATGATTTGGTTCCAGAA 3125

Qy      448 ---------------------------------------GlyGluProThrHisIleIle 454
Db     3126 GAAACTTGCACATCTAACAATCTGCCTTATGATGCAGATATGGAAGTTGCTGGTGCAGCT 3185

Qy      455 AspTyrLeuLysPheSer---IleAlaArgProAlaIleAspLysGluLeuGluAlaPhe 473
Db     3186 GATTTTCTCTCGAGTGCTTGGCAGTGCAAGTGCTCATATGAAACACAACTGAACGCACTG 3245

Qy      474 HisAsnAlaMetLysAlaAlaLysAspThrGlu---AspGlyAlaHisPheTrpAspPro 492
Db     3246 CTCAACCAATATGGCGTGCGCACTGAAGCAGAGCTTGTGACAGAGCATATATGG------ 3299

Qy      493 AspLeuAlaSerTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSerAla 512
Db     3300 TCGCTTCCCAAGTAC--------------------------AGCAGCAGAAGGCAGGGG 3332

Qy      513 LeuLeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeu 532
Db     3333 GACATAAAGGAGAGGTTGAAGAATGCATACTATGCTCTTCACAAGGAGTTT---AGGAGC 3389

Qy      533 ValLysAsnLysGluMetArgAspSerLysAspProTyrProValArgValAsnGlnVal 552
Db     3390 ATTTTCGAAAGCATTGTGACAGATCAAACAGAGATCTCTGATGATGAGAAAAGTCGGTTT 3449

Qy      553 TyrGlu-----------LysTrpCysAlaIleThr------ProGluAlaMetAspLys 566
Db     3450 TACGAGATGAAGGCCTCCGCTTGGTACCAGGTAACCTACCATCCTGAATGGGTCCAGAAG 3509

Qy      567 SerGlyAlaAsnTyrAspSer 573
Db     3510 TCAAGGGAAATGTTCAAGTCT 3530
```

```
RESULT 11
AAH77703
ID   AAH77703 standard; DNA; 6863 BP.
XX
AC   AAH77703;
XX
DT   13-NOV-2001  (first entry)
XX
DE   Nucleotide sequence of the SGS2 gene.
XX
KW   SGS2; RNA-dependent RNA polymerase; transgene silencing;
KW   transgene stability; crop plant; viral resistance; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   promoter        1. .850
FT                   /*tag=  a
FT   exon            851. .3564
FT                   /*tag=  b
FT                   /number= 1
FT   intron          3565. .3986
FT                   /*tag=  c
FT                   /number= 1
FT   exon            3987. .4862
FT                   /*tag=  d
FT                   /number= 2
FT   terminator      4863. .6863
FT                   /*tag=  e
XX
PN   FR2804128-A1.
XX
PD   27-JUL-2001.
XX
PF   26-JAN-2000; 2000FR-00001007.
XX
PR   26-JAN-2000; 2000FR-00001007.
XX
PA   (RHOB-) RHOBIO.
PA   (INRG ) INST NAT RECH AGRONOMIQUE.
XX
PI   Beclin C,  Elmayan T,  Mourrain P,  Vaucheret H;
XX
DR   WPI; 2001-543303/61.
XX
PT   A new plant SGS2 gene involved in encoding an RNA-dependent RNA
PT   polymerase and in transgene silencing, increases transgene stability and
PT   expression in transgenic plants when it is inactivated.
XX
PS   Claim 2; Page 30-32; 46pp; French.
XX
CC   The present sequence encodes a plant SGS2 polypeptide. SGS2 is an RNA-
CC   dependent RNA polymerase and is involved in transgene silencing.
CC   Inactivation of SGS2 is used to increase transgene stability and
CC   expression in plants, particularly crop plants, especially maize, corn,
CC   barley, sorghum, soya, sugar cane, beet, tobacco or cotton plants.
CC   Overexpression of SGS2 can be used to increase resistance to viral
CC   infection in plants
XX
SQ   Sequence 6863 BP; 1862 A; 1354 C; 1485 G; 2162 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.29e-10       Length:         6863
Score:                  220.50         Matches:        77
Percent Similarity:     41.96%         Conservative:   64
Best Local Similarity:  22.92%         Mismatches:     128
Query Match:            7.33%          Indels:         67
DB:                     4              Gaps:           11

US-09-913-878A-2_COPY_710_1282 (1-573) x AAH77703 (1-6863)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
                |||:::|||  ::::::     :::|||     :::   :::    |||   |||
Db       2608 GGAATTGGTACAATCACACCTGACCTCGCTGACGAAGTAATGGAGAAACTTAAGTTG--- 2664

Qy         21 AspVal---------ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrp 37
              ||||||         |||   |||   |||  |||:::         ||||||:::
Db       2665 GATGTGCACTACAGCCCTTGTGCTTATCAGATACGTTACGCAGGTTTCAAAGGGGTT--- 2721

Qy         38 ValIleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln--- 56
                                                      :::    :::||||||:::
Db       2722 -----------------------------------GTTGCTCGTTGGCCATCAAAAAGT 2745

Qy         57 --------------------ArgLysTrpGluCysAspPheValAspLysHisGlnArg 69
                                  |||             |||      ||||||
Db       2746 GATGGAATCAGGCTAGCCCTTCGAGACAGTATGAAGAAGTTCTTTTCCAAACAT---ACG 2802

Qy         70 ThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeu 89
                 ||||||:::   |||      :::      :::      ||||||   |||:::
Db       2803 ATCTTGGAGATCTGTTCC---TGGACGAGGTTTCAACCTGGGTTCTTAAATCGGCAGATA 2859

Qy         90 LeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAsp 109
              :::   :::|||                                      :::   |||
Db       2860 ATTACCCTTCTATCC-------------------------GTACTAGGTGTTCCGGAT 2892

Qy        110 ArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgPro 129
                 :::      |||:::|||                      :::    |||||||||
Db       2893 GAAATATTCTGGGATATGCAG---------GAATCCATGCTCTATAAACTGAACCGCATC 2943

Qy        130 ValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHis 149
              ::::::               :::|||      ::::::
Db       2944 CTTGATGATACAGATGTGGCATTTGAAGTTCTCACGGCATCATGTGCT------------ 2991

Qy        150 GlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeu 169
                                               |||       |||         :::
Db       2992 --------------------------------GAACAGGGAAACACTGCAGCTATCATG 3018

Qy        170 MetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeu 189
              :::::::::||||||   ||||||   :::   :::|||:::      :::      :::
Db       3019 CTTAGTGCAGGTTTCAAACCAAAAACCGAGCCGCATCTACGCGGGATGTTGTCTTCAGTC 3078

Qy        190 GlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAla 209
              :::      :::            |||:::   |||      |||   |||
Db       3079 AGAATTGCACAACTCTGGGGTCTCAGAGAAAAATCTCGTATTTTTGTTACTTCAGGAAGG 3138

Qy        210 TyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGly 229
              ::::::               |||      |||:::||||||      :::      :::
Db       3139 TGGCTAATGGGTTGCCTAGACGAAGCAGGGATACTTGAACATGGCCAATGCTTTATTCAA 3198

Qy        230 PheSer------------------------------------SerLysPheArgAspGlu 237
                 |||                                      |||:::|||::::::
Db       3199 GTCTCTAAACCGTCTATAGAAAATTGTTTCTCCAAACATGGGTCTCGTTTTAAGGAGACA 3258

Qy        238 GluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHis 257
              :::         ::::::               |||   :::|||::::::|||
Db       3259 AAGACAGATCTGGAAGTAGTTAAAGGCTAT---GTAGCCATTGCTAAGAATCCTTGTCTT 3315

Qy        258 PheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLys 277
                 |||   |||::::::   :::   ||||||      |||:::||||||   :::
Db       3316 CACCCAGGGGATGTAAGGATTTTAGAAGCTGTTGATGTACCCCAGCTGCATCACATGTAT 3375

Qy        278 AspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGly 297
              |||   :::||||||         |||||||||   |||            :::   ||||||
Db       3376 GACTGCCTTATTTTCCCTCAGAAAGGTGATAGGCCGCATACAAACGAAGCTTCTGGCAGT 3435

Qy        298 AspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleVal 313
              |||   |||||||||:::   :::|||   ||||||   :::::::::
Db       3436 GACCTTGACGGGGACCTGTACTTTGTGGCTTGGGATCAGAAACTCATC 3483

RESULT 12
AAC63738
ID   AAC63738 standard; cDNA; 1281 BP.
XX
AC   AAC63738;
XX
DT   07-FEB-2001  (first entry)
```

```
XX
DE   Maize RNA-directed RNA polymerase EST contig, SEQ ID NO: 3.
XX
KW   Maize; plant; RNA-directed RNA polymerase; expressed sequence tag; EST;
KW   gene mapping; gene marker; plant virus resistance; plant breeding; ss.
XX
OS   Zea mays.
XX
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT,  Orozco EM,  Weng Z,  Sakai H,  Cahoon RE,  Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   P-PSDB; AAB28529.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 2; Page 33; 62pp; English.
XX
CC   The present sequence is one of a number of cDNA molecules which encode
CC   plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC   probes for genetically and physically mapping genes, and as markers for
CC   traits linked to those genes. They are useful for controlling gene
CC   expression and provide mechanisms to engineer plant virus resistance.
CC   They are also useful for plant breeding to develop lines with desired
CC   phenotypes
XX
SQ   Sequence 1281 BP; 359 A; 282 C; 299 G; 338 T; 0 U; 3 Other;

Alignment Scores:
Pred. No.:              2.33e-11     Length:         1281
Score:                  218.00       Matches:        81
Percent Similarity:     39.53%       Conservative:   70
Best Local Similarity:  21.20%       Mismatches:     153
Query Match:            7.24%        Indels:         78
DB:                     3            Gaps:           11

US-09-913-878A-2_COPY_710_1282 (1-573) x AAC63738 (1-1281)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
               |||:::|||::::::|||          |||:::::::::     ::::::||||||
Db        113 GGTATCGGGAAGATTTCTACTAGATTTGCCAGACAAGTCGCCAAATTAATTGGCTTAGAC 172

Qy         21 AspVal------ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpVal 38
                          |||||||||   |||   |||:::|||        ||||||      |||
Db        173 CCAGCTCATCCTCCTTCTGCTTTTCAAATAAGGTATGGGGCTATAAAGGA------GTC 226

Qy         39 IleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLys 58
              |||   :::|||   |||            :::   :::       ||||||   :::|||
Db        227 ATCACTATTGACCCTACATCCTTTTTCAAT---CTTTCTCTGCGACCTAGTATGAAGAAG 283

Qy         59 TrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSer 78
              :::|||
Db        284 TTTGAATCG--------------------------------------------------- 292

Qy         79 GluLeuLysSerAlaGlyLeuAsnLeu------------GlnLeuLeuProValLeuGlu 94
                    ||||||        ||||||:::                   :::      |||
Db        293 ------AAGAGCACTATGCTGAACATTACAAATTGGAGTAAGTCTCAGCCATGTTATGTG 346

Qy         95 AspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAsp 114
              :::|||            :::               |||   |||   :::
Db        347 AACCGTGAAATTATCTCTCTTCTTTCAACATTGGGGATAAAGGATGAAGTATTTGANTCG 406

Qy        115 LeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGln 134
              :::|||:::            :::         :::                |||
Db        407 ATGCAACAAGATGACATGCACGAATCAGATGGAATGCTAACAAATAAAGAA--------- 457

Qy        135 TrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPhe 154
                                                                     :::
Db        458 ---------------------------------------------GCTGCTTTGTCTGTC 472

Qy        155 LeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPhe 174
              |||      :::                  :::|||         ::::::      |||:::
Db        473 CTAGGGAAAATTGGTGGCGGTGATACCAAGACGGCAGCTGATATGCTTCTTCAAGGCTAT 532

Qy        175 AspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCys 194
              :::|||      :::   ||||||      |||                :::      :::
Db        533 GAACCAAGTTCAGAGCCTTACCTATTAATGATTCTTAAAGCCCATCGGGCTAATAGGCTG 592

Qy        195 AspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIle 214
                 ::::::::::::      |||   |||   :::            :::
Db        593 ACCGACATAAGAACTCGGTGTAAGATTCATGTCCAGAAAGGCCGTGTTCTTATTGGTTGT 652

Qy        215 AlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPhe 234
                 |||         ||||||         :::||||::::::         :::
Db        653 TTGGATGAAACTTGCAAATTAGAGTATGGCCAAGTTTACATCAGAATTACAAAGAATCGC 712

Qy        235 ArgAspGlu---------GluGluSerPhe--------------------ThrLeuLeu 244
              :::::::::          |||:::   |||                     |||   :::
Db        713 AAGGAGCAGAAGTACAGTGAACAGCCGTTCTTTTGCAACGATGATGGCAAAACAGCTGTA 772

Qy        245 SerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArg 264
                       |||   :::   ::::::|||            |||   |||::::::
Db        773 ATTGTCGGAAAAGTTGCAATCACAAAAAACCCTTGTCTCCATCCTGGTGATGTCAGAGTA 832

Qy        265 ValArgAlaValPheLysProGluLeu------HisSerLeuLysAspValIleIlePhe 282
              :::   ||||||:::   |||   |||               |||   |||   ::::::|||
Db        833 CTTGAAGCTGTATATGACCCTGGATTGGATGCTAGGGGTCTTATTGATTGTGTTGTATTT 892

Qy        283 SerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAsp 302
                 :::|||:::   |||         :::   ||||||||||||   ||||||||
Db        893 CCTCAGAGAGGGGAAAGGCCTCATCCGAATGAATGCTCCGGGGGCGATTTGGATGGCGAC 952

Qy        303 MetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetPro 322
              :::   ::::::   ||||||   :::::::::            |||:::|||   |||
Db        953 CTCTTCTTTATTACTTGGGATGACAAACTGATTCCGGAGAAGGTTGATGCACCTATG--- 1009

Qy        323 LeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMet 342
                                                 |||   |||      :::      :::
Db       1010 ---------------------------------GACTACACTGCAACGAGGCCACGCATA 1036

Qy        343 AlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLys 362
                 |||                                |||   :::   |||||||||
Db       1037 ATGGACCATGCT-----------------------GTTACACTTGAGGAAATTCAGAAG 1072

Qy        363 SerPhe 364
                 |||
Db       1073 CACTTC 1078

RESULT 13
ABL27117
ID   ABL27117 standard; DNA; 4818 BP.
XX
AC   ABL27117;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 32824.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
```

```
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 32824; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4818 BP; 1368 A; 1145 C; 1214 G; 1091 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0.000852      Length:        4818
Score:                  148.50        Matches:       137
Percent Similarity:     34.38%        Conservative:  93
Best Local Similarity:  20.48%        Mismatches:    250
Query Match:            4.93%         Indels:        189
DB:                     4             Gaps:          33

US-09-913-878A-2_COPY_710_1282 (1-573) x ABL27117 (1-4818)

Qy          1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGly------ 18
Db         25 GGCAGCCGAAGGACTCGAAGATCCCTAGCTTCGGGTGGTGGTGATCTCATTGGCCCCTAT 84
Qy         19 -----------------LeuGlyAspValProSerAlaValGlnGlyArgPhe------ 30
Db         85 ATCTATTGCATCAACTTCTATGATAATGGATTGAGTGCGGTGCGGGATGAGTTTGTGAAG 144
Qy         31 -----------GlySerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAsp 46
Db        145 CGGCTTGAAAGTGGGCTACGACGTGCCATGTTCGTCATCGACTTGGATAAGGAAGTTGCG 204
Qy         47 GluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLys 66
Db        205 GAG---------GCCAATTATCAGTTCCAACGAAAG------------------------ 231
Qy         67 HisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGly---Leu 85
Db        232 ---TCGAGGACCAAGAAAGCTGGATCTCTG------GACATGAAGTCCTTAGATCCATTG 282
Qy         86 AsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGln 105
Db        283 AACATT---GTCCTGCAATGCATTCAGAACTGTGTGGAAGAGTTTGACAGAATAACGAAG 339
Qy        106 AlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAla 125
Db        340 CAGATTGAGTCAGAGGCGAATTTCGACCTGTACTCCTACTGGCAACGA------------ 387
Qy        126 LeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThr 145
Db        388 ---AATATCCCTGCACAGCTCAAGAACCTATTG--------------------------- 417
Qy        146 ArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluGluThr 165
Db        418 -----------------GCGCCTAAAAAAAAGGGTGCTCCAGATACGAAGCGTGAC--- 456
Qy        166 LeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIle 185
Db        457 CTAAATACAGCTAGATCC---------------AAAACGGGCAAGACATCGAGCGAC--- 498
Qy        186 AlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgVal 205
Db        499 AAGATGCGCCTGGAGAAGATCAAGTGCTCGTCCGTTACAGCCAAGGAGCACCACAAAGTT 558
Qy        206 GlyArgSerAla------------------TyrIle------------------------ 211
Db        559 GGCAAGGGAGCGCCGGTGTCCAAGCGAGAATACATTCGAGACAATCCGACGTTCACGAAA 618
Qy        212 --------------------TyrMetIleAlaAspPheTrpGlyValLeuGluGluAsn 224
Db        619 ATGCTGATCCTGATCGTGGGGCAAATGGATAACGACTTCTACAAATTGCTATTGGCCTAC 678
Qy        225 GluValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeu 244
Db        679 GAACAACCGCTGCGAGCTATTATTAACTTTATGCCGGAGGAGAGTGCCTAC--------- 729
Qy        245 SerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArg 264
Db        730 ---------GATATTTTAGTGCCCCGACCGCGACGGGAA------CAGCAGCTCCAGGAA 774
Qy        265 ValArgAlaValPheLysProGluLeuHisSerLeu------------------LysAsp 278
Db        775 GCGTTGGCTGCCATCCAGAAGGACATCCATTCGCTGCGAAAGAGGACTCCCACAAAACCG 834
Qy        279 ValIleIlePheSerThrLysGlyAspValPro---------------LeuAlaLysLys 293
Db        835 GTTGGCATCTTTCAGCAGCGG------CTTCCCCAGATGTCCGCTTGTATGGCAGCCAAA 888
Qy        294 LeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu------ 311
Db        889 TTCACGGACGATATCTTTGAC---CAGCTAAGTTGGTACATGTACGACGTGGAAACCCTG 945
Qy        312 --------------------IleValAspGlyPheValAsnAlaGluMetProLeuGlu 324
Db        946 AGGGAGTGGTATCGATCCTACTACGTGGATCCGTACAAGGAGACAAATATTAAAATGGAT 1005
Qy        325 ProAspLeuSer------------------------------------------------Arg 329
Db       1006 CCCTCGATGTCGTTGGTGGAAGCATTCCACATTGTCGAATCGCTGAGCATTCAGGGGCAC 1065
Qy        330 TyrLeuLys--------------------LysAspLysThrThr------------Phe 338
Db       1066 TATCTCAAGACGCAAATGCCGGCTTCCCGTAATGATAATGGTACCGTTTACCTATATCTG 1125
Qy        339 LysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAsp 358
Db       1126 GAGTCGCTAATGAGCACGTTTGGTAATCCAAAGGAACTGATGACCGAAACGGAGGTCCTG 1185
Qy        359 MetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsn 378
Db       1186 CTTCTGGCAAATCCCACGCCA---TCGGTTCAAACTCGAGTTTTGGACAAAGTTAAGGTA 1242
Qy        379 TyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIle--- 397
Db       1243 TACGCCAATTCGTTTAAGAGCATCGTAAAGATGCACAAAAACGACCGAATTTTTGTGGAG 1302
Qy        398 ---LeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsn 416
Db       1303 GAGGCCAATGGACTGCTGTCCAATCTGATTTCATATATGGATCAGAGTCTCATGCGAAAT 1362
Qy        417 Glu-----------AlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeu 432
```

```
Db       1363 GTCTACCACGGCTGTCTGGAGTATAATCTTTTGCGACGATACTTTACATCCGGGTATATC 1422

Qy        433 SerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHis 452
                       :::         |||          :::     |||        |||||     :::
Db       1423 ---------AACAGCTCACTGAAGTGTCAGCCCTACAATGGTGATGTGGAACCGATCTAT 1473

Qy        452 ------------------------------------------------------------ 452

Db       1474 CTGCCCAAGTTTGCCGAACTCTATCTCACGGATGACTCGGTTATGCAACGAATTATCGAG 1533

Qy        453 ---IleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGlu 471
                 :::   ||||||      ||||||:::                 :::       |||
Db       1534 CTGGTCAAGGACTACGATGACTTCAGCCTGGAGGAGGTGCAGCCCAATGTCCATCTCTAT 1593

Qy        472 AlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAla------------ 487
                 |||       ::::::               :::   |||:::
Db       1594 ACTTTTCGAAGGTCACTAAACGAGGTATTCGAACAGGAAAATCAAACAGTCATCCCAACA 1653

Qy        488 ------HisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGluIleSer 505
                       |||   |||        |||                |||   |||::::::
Db       1654 CGGCTCTGCTTCAGGGATTTTACACTGTACGAAATGGAGAATTTTCTAAAGGACTTG--- 1710

Qy        506 AspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluVal 525
                                                            |||      ||||||
Db       1711 --------------------------------GTCACCCCACAAAGGTTCAACGAGGTG 1737

Qy        526 GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAsp---------SerLys 542
                 :::   |||      |||   :::      |||      |||           ::::::
Db       1738 ACCGAAAGTTAT------TTGTCACGACAAACGGAGAGCGCTGATAATATTGCGACCCAA 1791

Qy        543 AspProTyrProValArgValAsnGln 551
              ||||||:::   ||||||   |||:::
Db       1792 GATCCATTCATAGTGAGGTGCAACGAG 1818
```

RESULT 14
ABL71773
ID   ABL71773 standard; cDNA; 282 BP.
XX
AC   ABL71773;
XX
DT   14-MAY-2002 (first entry)
XX
DE   Corn tassel-derived polynucleotide (cdps) SEQ ID NO:1147.
XX
KW   Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPs;
KW   inheritance; characteristic; growth; development; disease resistance;
KW   environmental adaptability; quality; yield; molecular marker;
KW   multigene trait; plant breeding; corn tassel; gene; ss.
XX
OS   Zea mays.
XX
PN   US2001051335-A1.
XX
PD   13-DEC-2001.
XX
PF   16-APR-1999; 99US-00294093.
XX
PR   21-APR-1998; 98US-0082567P.
XX
PA   (LALG/) LALGUDI R V.
PA   (ITOL/) ITO L Y.
PA   (SHER/) SHERMAN B K.
XX
PI   Lalgudi RV, Ito LY, Sherman BK;
XX
DR   WPI; 2002-163647/21.
XX
PT   Novel purified corn tassel-derived polynucleotide useful for determining
PT   altered gene expression, to recover regulatory elements and to follow
PT   inheritance of desirable characteristics through hybrid breeding
PT   programs.
XX
PS   Claim 1; SEQ ID NO 1147; 201pp; English.
XX
CC   The present sequence describes a purified corn tassel-derived
CC   polynucleotide sequence (cdps) comprising a nucleic acid sequence
CC   selected from those given in ABL70627 to ABL76833. The cdps sequences
CC   encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)
CC   can be used for determining altered gene expression, to recover
CC   regulatory elements and to follow inheritance of desirable
CC   characteristics through hybrid breeding programs. (I) are also useful in
CC   the evaluation, and alteration of desired characteristics associated with
CC   growth and development, disease resistance, environmental adaptability,
CC   quality and yield, and as molecular markers for studying inheritance of
CC   multigene traits in a plant breeding program. (I) can be used to produce
CC   a tassel-specific profile of gene transcription, a transcript image, to
CC   clone regulatory elements for use in transformation vectors, to express a
CC   polypeptide, to identify, isolate or extend identical or related corn
CC   tassel nucleic acid sequences from DNA libraries, in nucleic acid
CC   hybridisation or amplification technologies, as query sequences to
CC   determine homology of known sequences, as probe for use in Southern or
CC   Northern hybridisation, and to identify the presence of and/or to
CC   determine the degree of similarity between two (or more) nucleic acid
CC   sequences
XX
SQ   Sequence 282 BP; 78 A; 67 C; 70 G; 67 T; 0 U; 0 Other;

Alignment Scores:			
Pred. No.:	2.04e-05	Length:	282
Score:	148.00	Matches:	26
Percent Similarity:	64.62%	Conservative:	16
Best Local Similarity:	40.00%	Mismatches:	23
Query Match:	4.92%	Indels:	0
DB:	6	Gaps:	0

US-09-913-878A-2_COPY_710_1282 (1-573) x ABL71773 (1-282)

```
Qy        249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
              :::::::::|||::::::|||                  |||   |||::::::   :::   ||||||
Db         21 ATCGTAATGGCAAAGAATCCATGCCTTCATCCAGGGGATGTCCGAATCCTTGAAGCTGTT 80

Qy        269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288
                       ||||||||||||   |||   |||      ::::::|||         ||||||:::
Db         81 GATGTGCCTGAACTGCATCACCTTGTTGATTGCTTGGTCTTCCCCAAGAAAGGTGAGAGG 140

Qy        289 ProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrp 308
              |||   |||   :::   ||||||   |||   |||||||||:::      :::|||   |||
Db        141 CCGCACGCGAATGAAGCATCTGGGAGTGATCTTGATGGGGATCTATACTTCGTAACATGG 200

Qy        309 AspProGluIleVal 313
              |||      ::::::
Db        201 GATGAAAACCTTATA 215
```

RESULT 15
ABL71722
ID   ABL71722 standard; cDNA; 274 BP.
XX
AC   ABL71722;
XX
DT   14-MAY-2002 (first entry)
XX
DE   Corn tassel-derived polynucleotide (cdps) SEQ ID NO:1096.
XX
KW   Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPs;
KW   inheritance; characteristic; growth; development; disease resistance;
KW   environmental adaptability; quality; yield; molecular marker;
KW   multigene trait; plant breeding; corn tassel; gene; ss.
XX
OS   Zea mays.
XX
PN   US2001051335-A1.
XX
PD   13-DEC-2001.
XX
PF   16-APR-1999; 99US-00294093.
XX

```
PR   21-APR-1998;   98US-0082567P.
XX
PA   (LALG/) LALGUDI R V.
PA   (ITOL/) ITO L Y.
PA   (SHER/) SHERMAN B K.
XX
PI   Lalgudi RV,  Ito LY,  Sherman BK;
XX
DR   WPI; 2002-163647/21.
XX
PT   Novel purified corn tassel-derived polynucleotide useful for determining
PT   altered gene expression, to recover regulatory elements and to follow
PT   inheritance of desirable characteristics through hybrid breeding
PT   programs.
XX
PS   Claim 1; SEQ ID NO 1096; 201pp; English.
XX
CC   The present sequence describes a purified corn tassel-derived
CC   polynucleotide sequence (cdps) comprising a nucleic acid sequence
CC   selected from those given in ABL70627 to ABL76833. The cdps sequences
CC   encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)
CC   can be used for determining altered gene expression, to recover
CC   regulatory elements and to follow inheritance of desirable
CC   characteristics through hybrid breeding programs. (I) are also useful in
CC   the evaluation, and alteration of desired characteristics associated with
CC   growth and development, disease resistance, environmental adaptability,
CC   quality and yield, and as molecular markers for studying inheritance of
CC   multigene traits in a plant breeding program. (I) can be used to produce
CC   a tassel-specific profile of gene transcription, a transcript image, to
CC   clone regulatory elements for use in transformation vectors, to express a
CC   polypeptide, to identify, isolate or extend identical or related corn
CC   tassel nucleic acid sequences from DNA libraries, in nucleic acid
CC   hybridisation or amplification technologies, as query sequences to
CC   determine homology of known sequences, as probe for use in Southern or
CC   Northern hybridisation, and to identify the presence of and/or to
CC   determine the degree of similarity between two (or more) nucleic acid
CC   sequences
XX
SQ   Sequence 274 BP; 68 A; 58 C; 72 G; 76 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0.000715      Length:        274
Score:                  132.00        Matches:       25
Percent Similarity:     58.67%        Conservative:  19
Best Local Similarity:  33.33%        Mismatches:    29
Query Match:            4.39%         Indels:        2
DB:                     6             Gaps:          1

US-09-913-878A-2_COPY_710_1282 (1-573) x ABL71722 (1-274)

Qy        249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
              |||   :::   ::::::|||         |||  |||::::::   :::   |||||
Db          2 GTTGCAATCACAAAAAACCCTTGTCTCCATCCTGGTGATGTCAGAGTACTTGAAGCTGTA 61

Qy        269 PheLysProGluLeu------HisSerLeuLysAspValIleIlePheSerThrLysGly 286
              :::   |||   |||            |||   |||   ::::::|||      :::|||
Db         62 TATGACCCTGGATTGGATGCTAGGGGTCTTATTGATTGTGTTGTATTTCCTCAGAGAGGG 121

Qy        287 AspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpVal 306
              :::   |||         :::   ||||||||||||   |||||||||:::   ::::::
Db        122 GAAAGGCCTCATCCGAATGAATGCTCCGGGGGCGATTTGGATGGCGACCTCTTCTTTATT 181

Qy        307 CysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet 321
                 ||||||   :::::::::         |||:::|||   |||
Db        182 ACTTGGGATGACAAACTGATTCCGGAGAAGGTTGATGCACCTATG 226

Search completed: March 31, 2004, 07:26:40
Job time : 385.609 secs
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:         March 31, 2004, 06:31:56 ; Search time 65.8587 Seconds
                                            (without alignments)
                                            4828.317 Million cell updates/sec

Title:          US-09-913-878A-2_COPY_710_1282
Perfect score:  3010
Sequence:       1 GVGRMSRSVAKRIRDVLGLG..........EKWCAITPEAMDKSGANYDS 573

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:      1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09913878/runat_29032004_082009_8395/app_query.fasta_1.2254
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878_@CGN_1_1_82_@runat_29032004_082009_8395 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	262	8.7	3731	3	US-08-811-583-1	Sequence 1, Appli
	2	124	4.1	274	4	US-09-313-294A-5730	Sequence 5730, Ap
	3	117.5	3.9	13977	4	US-09-484-970B-60	Sequence 60, Appl
	4	114	3.8	1491	4	US-09-543-681A-2293	Sequence 2293, Ap
	5	113	3.8	8494	4	US-08-961-527-163	Sequence 163, App
	6	112.5	3.7	6702	3	US-08-987-439-2	Sequence 2, Appli
	7	111	3.7	12311	3	US-08-750-717-1	Sequence 1, Appli
	8	110.5	3.7	6909	2	US-08-804-196-1	Sequence 1, Appli
	9	110.5	3.7	6909	2	US-08-658-340-1	Sequence 1, Appli
	10	110.5	3.7	6909	3	US-08-746-111-26	Sequence 26, Appl
	11	110.5	3.7	6909	4	US-08-454-353A-1	Sequence 1, Appli
	12	110.5	3.7	6909	4	US-09-165-019-1	Sequence 1, Appli
	13	107.5	3.6	2077	4	US-09-620-312D-985	Sequence 985, App
	14	107.5	3.6	5307	4	US-09-489-039A-4533	Sequence 4533, Ap
	15	105.5	3.5	15462	3	US-09-073-492-1	Sequence 1, Appli
	16	104	3.5	3648	4	US-09-134-001C-2482	Sequence 2482, Ap
	17	103.5	3.4	6960	2	US-08-841-349-3	Sequence 3, Appli
	18	103.5	3.4	6960	4	US-09-431-184A-3	Sequence 3, Appli
	19	103.5	3.4	8176	2	US-08-841-349-5	Sequence 5, Appli
	20	103.5	3.4	8176	4	US-09-431-184A-5	Sequence 5, Appli
	21	103	3.4	1724	4	US-09-117-860-36	Sequence 36, Appl
	22	103	3.4	4495	4	US-09-620-312D-352	Sequence 352, App
	23	103	3.4	10320	4	US-09-091-501B-9	Sequence 9, Appli
c	24	102	3.4	1263	4	US-09-252-991A-2381	Sequence 2381, Ap
	25	102	3.4	2814	4	US-09-252-991A-2537	Sequence 2537, Ap
	26	101.5	3.4	7881	2	US-08-751-189-1	Sequence 1, Appli
	27	101.5	3.4	7881	2	US-09-060-836-1	Sequence 1, Appli
	28	101.5	3.4	7881	3	US-09-184-445-1	Sequence 1, Appli
	29	101	3.4	12284	2	US-08-876-991-1	Sequence 1, Appli
	30	101	3.4	12284	2	US-09-059-853-1	Sequence 1, Appli
	31	100.5	3.3	8351	1	US-08-198-446B-14	Sequence 14, Appl
	32	100.5	3.3	8351	2	US-08-870-693-14	Sequence 14, Appl
c	33	100.5	3.3	19307	3	US-08-836-022A-10	Sequence 10, Appl
c	34	100.5	3.3	19307	3	US-09-427-048A-10	Sequence 10, Appl
	35	100	3.3	7453	4	US-09-620-312D-248	Sequence 248, App
	36	100	3.3	7501	4	US-09-620-312D-249	Sequence 249, App
	37	100	3.3	7596	4	US-09-023-655-1463	Sequence 1463, Ap
	38	100	3.3	11283	2	US-08-603-753D-3	Sequence 3, Appli
	39	100	3.3	11283	3	US-09-099-753-3	Sequence 3, Appli
	40	100	3.3	11283	3	US-08-986-106-3	Sequence 3, Appli
	41	99.5	3.3	3159	4	US-09-134-000C-3215	Sequence 3215, Ap
c	42	99.5	3.3	4702	4	US-08-956-171E-268	Sequence 268, App
	43	99.5	3.3	5923	4	US-09-976-594-907	Sequence 907, App
	44	99.5	3.3	5952	4	US-09-687-875A-1	Sequence 1, Appli
	45	99.5	3.3	7886	2	US-08-751-189-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-08-811-583-1
; Sequence 1, Application US/08811583
; Patent No. 6218142
;  GENERAL INFORMATION:
;    APPLICANT:  Wassenegger, Michael
;    APPLICANT:  Riedel, Leonhard
;    APPLICANT:  Schiebel, Winfried
;    APPLICANT:  Sanger, Heinz
;    TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING
;    TITLE OF INVENTION:  POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
;    TITLE OF INVENTION:  RNA-DIRECTED RNA POLYMERASE (RdRP)
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & NEAVE
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/811,583
;      FILING DATE:  05-MAR-1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley, James F.
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  MPG-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-596-9000
;      TELEFAX:  212-596-9090
```

```
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3731 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Tomato
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  194..3535
US-08-811-583-1

Alignment Scores:
Pred. No.:              2.26e-20      Length:        3731
Score:                  262.00        Matches:       127
Percent Similarity:     37.66%        Conservative:  85
Best Local Similarity:  22.56%        Mismatches:    222
Query Match:            8.70%         Indels:        130
DB:                     3             Gaps:          17

US-09-913-878A-2_COPY_710_1282 (1-573) x US-08-811-583-1 (1-3731)

Qy        1 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 20
Db     1772 GGAATTGGTAAAATATCTGGTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCTTCAA 1831

Qy       21 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 40
Db     1832 TATACCCCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGGTGTTGAT 1891

Qy       41 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 60
Db     1892 CCGGATTCATCAATGAAG---------TTGTCTTTGAGAAAGAGCATGTCGAAATATGAA 1942

Qy       61 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 80
Db     1943 TCAGACAACATA---------------AAGTTAGATGTCCTTGGA---TGGAGCAAATAT 1984

Qy       81 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 100
Db     1985 CAGCCTTGTTATCTTAATCGTCAACTGATTACGCTCTTGTCT------------------ 2026

Qy      101 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 120
Db     2027 ---------ACACTTGGAGTGAAAGATGAAGTTCTCGAACAGAAGCAAAAGGAAGCTGTA 2077

Qy      121 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
Db     2078 GATCAGCTTGATGCTATCTTGCATGATTCTTTGAAGGCACAGGAGGCTTTGGAATTGATG 2137

Qy      141 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 160
Db     2138 TCT---------------------------------------------------CCTGGA 2146

Qy      161 SerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLys 180
Db     2147 GAGAACACTAATATTCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCC 2206

Qy      181 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 200
Db     2207 TTTCTTTCAATGATGTTGCAAACCTTCCGCGCATCCAAGTTGCTCGATTTGCGGACTAGA 2266

Qy      201 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 220
Db     2267 TCAAGAATATTTATTCCAAATGGAAGAACAATGATGGGATGTTTGGATGAATCCAGAACC 2326

Qy      221 LeuGluGluAsnGluValHisValGlyPheSer------------SerLysPheArgAsp 236
Db     2327 TTGGAATATGGTCAGGTGTTTGTTCAGTTTACTGGTGCTGGACATGGAGAGTTTTCTGAC 2386

Qy      237 GluGluGluSerPhe------------------------ThrLeuLeuSerAspCysAsp 248
Db     2387 GATTTACATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAAT 2446

Qy      249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
Db     2447 GTGGTTGTTGCAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTTAAAGGCTGTA 2506

Qy      269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288
Db     2507 AATGTTCGAGCGCTGCACCACATGGTAGATTGTGTTGTATTCCCTCAGAAAGGAAAAAGA 2566

Qy      289 ProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrp 308
Db     2567 CCTCATCCGAATGAATGTTCTGGGAGTGATTTGGATGGGGATATCTACTTTGTTTGCTGG 2626

Qy      309 AspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSer 328
Db     2627 GATCAAGACATGATCCCGCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCCAGC 2686

Qy      329 ArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGly 348
Db     2687 ATACAGTTGGACCATGAT------------------------------------------ 2704

Qy      349 SerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeu 368
Db     2705 -----------------GTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATT 2746

Qy      369 GlnProAsnPheLeuGlyMetCysThrAsn------------TyrLysGluArgLeuCys 384
Db     2747 GTGAATGACAGTTTGGGAATCATAGCAAATGCCCATGTCGTATTTGCAGACAGA------ 2800

Qy      385 TyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsn 404
Db     2801 ---GAACCTGATATGGCCATGAGTGATCCATGCAAAAAACTTGCTGAGCTCTTTTCAATT 2857

Qy      405 LeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArg 424
Db     2858 GCAGTGGACTTTCCAAAGACTGGTGTTCCCGCTGAAATACCATCT------CAGTTGCGC 2911

Qy      425 ArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTr 444
Db     2912 ------------------------------------CCTAAAGAATACCCAGACTTCATG 2935

Qy      444 pLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgPr 464
Db     2936 GATAAGCCGG----------------------------------------ACAAGACC 2953

Qy      464 oAlaIleAspLysGlu------LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAs 482
Db     2954 AGCTATATCTCAGAAAGAGTTATTGGAAAGCTTTTC-AGGAAAGTGAAG----------- 3001

Qy      482 pThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLy 502
Db     3002 -------GACAAAGCA------------CCTCAGGCTAGCTCTATCGCGACCTTCACAAG 3042

Qy      502 sGluIleSerAspLysSerArgSerSerAlaLeu------------------------- 513
Db     3043 AGATGTTGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACAT 3102

Qy      514 ----------------LeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGl 528
Db     3103 TGACGAAGCTTTTGACTACAAAACTGAATATGACAACAAGCTGGGTAATTTAATGGACTA 3162

Qy      528 uTyrGly 530
Db     3163 CTATGGC 3169

RESULT 2
US-09-313-294A-5730
; Sequence 5730, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT:  Lalgudi, Raghunath V.
```

```
;  APPLICANT:  Ito, Laura Y.
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE:  PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  7600
;  SOFTWARE:  PERL Program
; SEQ ID NO 5730
;   LENGTH: 274
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700350587H1
;   NAME/KEY: unsure
;   LOCATION: 41, 232, 244, 248, 257-258, 268-269
;   OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5730

Alignment Scores:
Pred. No.:               8.6e-06       Length:         274
Score:                   124.00        Matches:        24
Percent Similarity:      56.16%        Conservative:   17
Best Local Similarity:   32.88%        Mismatches:     30
Query Match:             4.12%         Indels:         2
DB:                      4             Gaps:           1

US-09-913-878A-2_COPY_710_1282 (1-573) x US-09-313-294A-5730 (1-274)

Qy        249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
              |||   :::   :::   |||         |||   |||::::::  :::   ||||||
Db         26 GTTGCAATCACAAAANACCCTTGTCTCCATCCTGGTGATGTCAGAGTACTTGAAGCTGTA 85

Qy        269 PheLysProGluLeu------HisSerLeuLysAspValIleIlePheSerThrLysGly 286
              :::   |||   |||            |||   |||   ::::::|||       :::|||
Db         86 TATGACCCTGGATTGGATGCTAGGGGTCTTATTGATTGTGTTGTATTTCCTCAGAGAGGG 145

Qy        287 AspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpVal 306
              :::   |||         :::   ||||||||||||   |||||||||:::   ::::::
Db        146 GAAAGGCCTCATCCGAATGAATGCTCCGGGGGCGATTTGGATGGCGACCTCTTCTTTATT 205

Qy        307 CysTrpAspProGluIleValAspGlyPheValAsnAla 319
                 ||||||   ::::::         |||:::|||
Db        206 ACTTGGGATGACAAACTGTTTCCGGANAAGGTTGATGCN 244

RESULT 3
US-09-484-970B-60
; Sequence 60, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
;  APPLICANT:  Jones, Karen A.
;  APPLICANT:  Volkmuth, Wayne
;  APPLICANT:  Walker, Michael G.
;  TITLE OF INVENTION:  BONE REMODELING GENES
;  FILE REFERENCE:  PB-0014 US
;  CURRENT APPLICATION NUMBER: US/09/484,970B
;  CURRENT FILING DATE:  2000-01-18
;  NUMBER OF SEQ ID NOS:  172
;  SOFTWARE:  PERL Program
; SEQ ID NO 60
;   LENGTH: 13977
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1
;   NAME/KEY: unsure
;   LOCATION: 11721-11761, 12294, 13969
;   OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-60

Alignment Scores:
Pred. No.:               0.0316        Length:         13977
Score:                   117.50        Matches:        130
Percent Similarity:      32.36%        Conservative:   102
Best Local Similarity:   18.13%        Mismatches:     254
Query Match:             3.90%         Indels:         231
DB:                      4             Gaps:           33

US-09-913-878A-2_COPY_710_1282 (1-573) x US-09-484-970B-60 (1-13977)

Qy         24 SerAlaValGlnGlyArgPheGlySer------AlaLysGlyMetTrpValIleAspVal 41
              :::|||   |||||||||   |||:::               |||         :::|||
Db       1376 ACAGCCCATCAGGGCCGGGTTGGTAATATTCTACAATTGGGAAGTAAGCTGATTGGAACA 1435

Qy         42 AspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArg----------- 57
                          :::::::::   |||
Db       1436 GGAAAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCA 1495

Qy         58 LysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAla 77
              :::|||||||||      |||         ::::::         :::         |||
Db       1496 AGATGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTTA 1555

Qy         78 SerGluLeuLysSerAlaGlyLeuAsn------LeuGlnLeuLeuProValLeuGluAsp 95
                 :::|||::::::   |||                      |||      |||:::
Db       1556 ATGGATCTCCAGAATCAGAAACTGAAAGAGTTGAATGACTGGCTAACAAAAACAGAAGAA 1615

Qy         96 ArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeu 115
              |||   |||         ||||||   :::                  :::   ||||||
Db       1616 AGAACAAGG---------AAAATGGAGGAAGAGCCTCTTGGACCTGATCTTGAAGACCTA 1666

Qy        116 GlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGln--- 134
              :::||||||      :::   |||         |||         :::|||      |||
Db       1667 AAACGCCAAGTACAACAACATAAG---GTGCTTCAAGAAGATCTAGAACAAGAACAAGTC 1723

Qy        135 ----------------------------TrpValTyrGluSerTyrSerSerArgAla 144
                                          |||   ||||||            |||
Db       1724 AGGGTCAATTCTCTCACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCACGCA 1783

Qy        145 ThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluGlu 164
              |||                                             :::   ||||||
Db       1784 ACTGCT-----------------------------------------GCTTTGGAAGAA 1801

Qy        165 ThrLeuAsnPheLeuMetAsn---------------SerGlyPheAspProLysLysGln 179
                 |||      |||   :::                        :::      :::
Db       1802 CAACTTAAGGTATTGGGAGATCGATGGGCAAACATCTGTAGATGGACAGAAGACCGCTGG 1861

Qy        180 LysTyrLeuGlnAspIle------AlaTrpAsp---LeuGlnLysArgLysCysAspThr 196
                 ||||||||||||            |||         |||   :::   :::|||
Db       1862 GTTCTTTTACAAGACATCCTTCTCAAATGGCAACGTCTTACTGAAGAACAGTGC------ 1915

Qy        197 LeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAsp 216
                                                                       :::
Db       1916 --------------------------------------------------CTTTTTAGT 1924

Qy        217 PheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAsp 236
                 |||         :::|||:::   |||:::            ::::::   |||:::|||
Db       1925 GCATGGCTTTCAGAAAAAGAAGATGCAGTGAACAAGATTCACACAACTGGCTTTAAAGAT 1984

Qy        237 GluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAla 256
              :::   |||      :::   |||                  ||||||   |||
Db       1985 CAAAATGAAATGTTATCAAGTCTTCAAAAACTGGCCGTTTTAAAAGCG------------ 2032

Qy        257 HisPheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeu 276
                          |||:::::::::   :::                  :::|||:::||||||
Db       2033 ------------GATCTAGAAAAGAAAAAGCAATCCATG---GGCAAACTGTATTCACTC 2077

Qy        277 LysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGly 296
              |||         ::::::   ||||||      :::      :::   :::|||
Db       2078 AAACAAGATCTTCTT---TCAACACTGAAGAATAAGTCAGTGACCCAGAAGACGGAA--- 2131

Qy        297 GlyAspTyrAspGlyAspMetAlaTrpVal---------------CysTrpAsp------ 309
                                 ||||||:::                  |||||||||
Db       2132 ---------------------GCATGGCTGGATAACTTTGCCCGGTGTTGGGATAATTTA 2170
```

```
Qy        310 ------------ProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluPro 325
Db       2171 GTCCAAAAACTTGAAAAGAGTACAGCACAGATTTCACAGGCTGTCACCACCACTCAGCCA 2230

Qy        326 AspLeuSerArgTyrLeuLysLysAspLys-----------ThrThrPheLysGlnLeu 341
Db       2231 TCACTAACACAGACAACTGTAATGGAAACAGTAACTACGGTGACCACAAGGGAACAGATC 2290

Qy        342 MetAlaSerHisGlyThrGlySer-----------------AlaAlaLysGluGlnThr 355
Db       2291 CTGGTAAAGCATGCTCAAGAGGAACTTCCACCACCACCTCCCCAAAAGAAGAGGCAGATT 2350

Qy        356 ThrTyrAsp---------------------------------------------Met 359
Db       2351 ACTGTGGATTCTGAAATTAGGAAAAGGTTGGATGTTGATATAACTGAACTTCACAGCTGG 2410

Qy        360 IleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeu--------------Gly 374
Db       2411 ATTACTCGCTCAGAAGCTGTGTTGCAGAGTCCTGAATTTGCAATCTTTCGGAAGGAAGGC 2470

Qy        375 MetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLys--- 393
Db       2471 AACTTCTCAGACTTAAAAGAAAAAGTCAATGCCATAGAGCGAGAAAAAGCTGAGAAGTTC 2530

Qy        394 ---------------ProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGln 408
Db       2531 AGAAAACTGCAAGATGCCAGCAGATCAGCTCAGGCCCTGGTGGAACAGATGGTGAATGAG 2590

Qy        409 SerLysGlnGlyIleValPheAsnGluAlaSer-------------------------- 419
Db       2591 GGTGTTAATGCAGATAGCATCAAACAAGCCTCAGAACAACTGAACAGCCGGTGGATCGAA 2650

Qy        420 ------------------------------TrpAlaGlnLeuArgArgGluLeuLeu--- 428
Db       2651 TTCTGCCAGTTGCTAAGTGAGAGACTTAACTGGCTGGAGTATCAGAACAACATCATCGCT 2710

Qy        429 ------GlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAspSerTrpLeu--- 445
Db       2711 TTCTATAATCAGCTACAACAATTGGAGCAGATGACAACTACTGCTGAAAACTGGTTGAAA 2770

Qy        446 ------------GlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIle 461
Db       2771 ATCCAACCCACCACCCCATCAGAGCCAACAGCAATTAAAAGTCAGTTAAAA------ATT 2824

Qy        462 AlaArgProAlaIleAspLys------------------GluLeuGluAlaPheHisAsn 475
Db       2825 TGTAAGGATGAAGTCAACCGGCTATCAGGTCTTCAACCTCAAATTGAACGATTAAAAATT 2884

Qy        476 AlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAla 495
Db       2885 CAAAGCATAGCCCTGAAAGAGAAAGGACAAGGACCCATGTTCCTGGATGCAGACTTTGTG 2944

Qy        496 SerTyrTyrThrPhePheLysGluIle---SerAsp---------LysSerArgSerSer 511
Db       2945 GCCTTTACAAATCATTTTAAGCAAGTCTTTTCTGATGTGCAGGCCAGAGAGAAAGAGCTA 3004

Qy        512 AlaLeuLeuPheThrThrLeu----------------------------------------- 518
Db       3005 CAGACAATTTTTGACACTTTGCCACCAATGCGCTATCAGGAGACCATGAGTGCCATCAGG 3064

Qy        518 ------------------------------------------------------------ 518
Db       3065 ACATGGGTCCAGCAGTCAGAAACCAAACTCTCCATACCTCAACTTAGTGTCACCGACTAT 3124

Qy        519 ---------LysAsnArgIleGlyGluVal-----------------------GluLys 527
Db       3125 GAAATCATGGAGCAGAGACTCGGGGAATTGCAGGCTTTACAAAGTTCTCTGCAAGAGCAA 3184

Qy        528 GluTyrGlyArgLeuValLysAsnLysGluMetArgAspSerLysAspProTyrProVal 547
Db       3185 CAAAGTGGCCTATACTATCTCAGCACCACTGTGAAAGAGATGTCGAAGAAAGCGCCCTCT 3244

Qy        548 ArgValAsnGlnValTyrGlu-----------------------LysTrp--------- 556
Db       3245 GAAATTAGCCGGAAATATCAATCAGAATTTGAAGAAATTGAGGGACGCTGGAAGAAGCTC 3304

Qy        557 --------------------CysAlaIleThrProGluAlaMetAspLys 566
Db       3305 TCCTCCCAGCTGGTTGAGCATTGTCAAAAGCTAGAGGAGCAAATGAATAAA 3355
```

RESULT 4
US-09-543-681A-2293
; Sequence 2293, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRAB:
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2293
;   LENGTH: 1491
;   TYPE: DNA
;   ORGANISM: Proteus mirabilis
US-09-543-681A-2293

Alignment Scores:

Pred. No.:	0.00213	Length:	1491
Score:	114.00	Matches:	88
Percent Similarity:	38.18%	Conservative:	80
Best Local Similarity:	20.00%	Mismatches:	152
Query Match:	3.79%	Indels:	120
DB:	4	Gaps:	22

US-09-913-878A-2_COPY_710_1282 (1-573) x US-09-543-681A-2293 (1-1491)

```
Qy        168 PheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrp 187
Db        223 TACATACTTGAAAAAAATAAAAAAAAAGATAGGCTGAAGAAATTAAAAGATATT------ 276

Qy        188 AspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArg 207
Db        277 ---CTTGAAAAAGAAAGA---------TTAACTGATAAAATAAAAATACTTGATAATAAG 324

Qy        208 SerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHis 227
Db        325 ---------------GTAATTGTATCA------GGAATAAAAGATGCTAATGATATAAAT 363

Qy        228 ValGlyPheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAsp--- 246
Db        364 GTATTATTTTCCTCTTTGTTAAAGAATAAAGATGATTCTAGTACACTTTTTAATAAAATA 423

Qy        247 ---------------------CysAspValLeuValAlaArgSerProAlaHisPhePro 259
Db        424 TTTAGTAATAAAAATACTATAATTAATAGTTTTATCTCTAAAAAAGAAAAAAAGATTTCT 483

Qy        260 SerAspIleGlnArgVal------ArgAlaValPheLysProGlu--------------- 272
Db        484 GATGACATAGAAAATATTGCAAATAAAATTATGCGTAAACCTAAATTTACATTATATAAT 543

Qy        273 ---LeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAla 291
Db        544 AATGTAAATAAATATCAAAATATAATTACTATTTCGAATAAA------------ATTTTT 591

Qy        292 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 311
Db        592 AAGGAACCAGAGAATACTGATATAACAAAAGAT------------------GATTCTCAT 633

Qy        312 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu 331
Db        634 AATAAAAATAATTTTGAAAAAACAGAA---------AAAACAGACCTCTCAAAAAAAACC 684
```

```
Qy        332 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 351
              ||||||    |||                                    :::|||
Db        685 AAAAAATCAAAAAAA-----------------------------------GCAGCCCTG 708

Qy        352 LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 371
                 ||||||         |||      ::::::    :::    |||       :::
Db        709 ATGGAACAACTTGATAAAGAT------AAACAGAAAATGAATGAGTTGTTAGCTACTTCT 762

Qy        372 PheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSer 391
                    |||:::             :::       ||||||:::|||    :::|||
Db        763 AGCCAAGGGTTATTAAAAACTGTAGAG-----------TATATAGATAAT---CTATCT 807

Qy        392 AsnLysProAlaIleIleLeuSerSerLeuValGlyAsn------------------Leu 405
              |||          ::::::||||||          |||                   |||
Db        808 AAT-----------TTAATTAGTTCTAAAACTGATAATGGTACTTGTAATTCCGACCTC 855

Qy        406 ValAspGlnSerLysGlnGlyIle-----------------------ValPheAsnGlu 417
              :::||||||   |||       |||                     |||||||||:::
Db        856 ATTGACCAATTCAAAATGTCTATTGAAACTTTGGTAAAAAAATGTAATGTCTTTAACAAA 915

Qy        418 AlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspPro 437
                 |||         ::::::::::::          :::   |||   |||
Db        916 GATAGTCAGTTTTTATCTAAAAAAGATATT---------GAGATAGAGTTAAGTGAC--- 963

Qy        438 MetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeu 457
                                                                       :::
Db        964 ---------------------------------------------------------ATA 966

Qy        458 LysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMet 477
              |||   :::|||                   :::||||||                |||
Db        967 AAAATAACAATT-----------------GAAGAACTTAATGAAATTATAAAAAAGATG 1008

Qy        478 LysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyr 497
                       |||         :::         :::   :::   |||   :::::::::
Db       1009 GATGGTTTTAAATCTAATCAGTTATATTGTTTTTATAAGGAAGCAGATAAATCTAATTTT 1068

Qy        498 TyrThr-------------------------------PhePheLysGluIleSerAsp 506
                 |||                               ||||||   |||   |||
Db       1069 ACTACAGTTAACGAGATTAATTTTAATTTTAAAAAGGTTTTTTTAATAATATAAAAGAT 1128

Qy        507 LysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluValGlu 526
                             :::               :::||||||          |||
Db       1129 ATGTTTTCTCAATTATCTTCTTATAATAAATCATTAAAGGAAAATAAAGGTTTTTTCCTT 1188

Qy        527 LysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSerLys---AspProTyr 545
              |||   |||      |||   |||||||||   :::|||      :::      :::
Db       1189 AAACTATATAAATTATTATTTAAAAACAAATACCTAAGAAAAAAAGAGATAAATGATAAT 1248

Qy        546 ProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrProGluAlaMetAsp 565
                 |||:::      ::::::   ::::::         :::      |||::::::|||
Db       1249 GCTGTTAAAAAATTAGAAATATTGGATAGGAATTTATTCCTTTTGCTTGAATCTCTAGAT 1308
```

RESULT 5
US-08-961-527-163
; Sequence 163, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 163:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8494 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-163

```
Alignment Scores:
Pred. No.:               0.0481      Length:         8494
Score:                   113.00      Matches:        125
Percent Similarity:      35.13%      Conservative:   77
Best Local Similarity:   21.74%      Mismatches:     216
Query Match:             3.75%       Indels:         157
DB:                      4           Gaps:           28

US-09-913-878A-2_COPY_710_1282 (1-573) x US-08-961-527-163 (1-8494)

Qy         12 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGly 31
              ||||||   |||::::::         ||||||:::
Db       3555 CGTATCTACGACTTGGTTAAATCTGGTGATTTA------------------GCTATTCAT 3596

Qy         32 SerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGlu 51
                 ||||||         |||   :::   |||   |||   |||         ::::::|||
Db       3597 AAAGCCAAGACATTTGTTGTTGATGAAGCAGATATGACCTTGGATATGGGATTCTTGGAA 3656

Qy         52 ThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 71
              |||                                   |||||||||         :::|||
Db       3657 ACT--------------------------------GTTGATAAGATTGCTGGCAGTCTT 3683

Qy         72 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuPro 91
                    :::   :::         :::   ||||||   :::      :::|||   |||
Db       3684 CCA---AAAGACTTGCAATTCATGGTCTTCTCAGCGACTATCCCACAAAAACTGCAACCA 3740

Qy         92 ValLeuGlu------------AspArgAlaArgAspLysValLysMetArgGlnAlaIle 107
              |||:::         :::         :::|||:::|||   :::         |||
Db       3741 TTCTTGAAAAAATACTTATCAAATCCTGTTATGGAGAAAATTAAGACCAAAACGGTTATT 3800

Qy        108 GlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAla------ 125
                 |||   :::   |||   |||                     |||:::|||
Db       3801 TCTGACACCATTGATAATTGGTTGATTTCGACCAAGGGACATGATAAGAATGCTCAAATT 3860

Qy        126 ---------------LeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 140
                             |||   :::|||      :::         ::::::   :::
Db       3861 TACCAGTTGACTCAGTTGATGCAGCCG------TATTTGGCAATGATTTTTGTTAACACT 3914

Qy        141 SerSerArgAlaThrArgValSer------------------------------------ 148
                 :::||||||      :::
Db       3915 AAAACGCGTGCTGATGAATTGCATTCATATCTGACTGCTCAAGGCTTGAAGGTTGCAAAA 3974

Qy        149 ---HisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsn 167
                 ||||||   :::                  |||      ::::::      :::|||
Db       3975 ATCCATGGCGATATT---------------GCCCCTCGTGAACGCAAGCGAATCATGAAT 4019

Qy        168 PheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrp 187
                 :::   |||      |||:::               |||:::         :::|||
Db       4020 CAGGTGCAAAATCTGGATTTTGAG---------------TATATT------GTCGCAACA 4058

Qy        188 AspLeuGlnLysArgLysCys-----------------------------AspThrLeu 197
              ||||||      |||                                   |||   :::
```

```
Db       4059 GATTTGGCAGCGCGTGGGATTGACATTGAAGGTGTCAGCCATGTCATCAATGATGCCATT 4118

Qy        198 LysSerLysLeuAsnIle---------ArgValGlyArgSerAla--------------- 209
                         |||:::            ||||||||||||:::
Db       4119 CCGCAAGACTTATCTTTTTTTGTTCATCGTGTTGGTCGTACTGGACGAAATGGCCTACCA 4178

Qy        210 ------------TyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGlu 225
                             :::|||      :::|||          :::   |||      :::
Db       4179 GGTACAGCTATTACCCTTTATCAGCCAAGTGATGACTCGGATATCCGTGAGTTGGAGAAA 4238

Qy        226 ValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeuSer 245
              :::   :::   ||||||   |||          :::      |||   |||
Db       4239 TTGGGAATCAAGTTTAGTCCTAAGATGGTCAAAGACGGGGAA---TTTCAAGATACCTAT 4295

Qy        246 AspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgVal 265
              |||   |||               |||               ||||||:::      :::
Db       4296 GACCGTGATCGTCGTGCCAACCGTGAGAAAAAACAAGATAAACTTGATATCGAAATGATT 4355

Qy        266 ArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLys 285
                       |||   |||   :::
Db       4356 GGTTTGGTTAAAAAGAAAAAG--------------------------------------- 4376

Qy        286 GlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrp 305
                                 ||||||:::      |||   |||         :::   |||
Db       4377 ------------------AAAAAAGTCAAACCGGGT---TATAAGAAGAAAATTCAATGG 4415

Qy        306 ValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluPro 325
                                    ||||||
Db       4416 GCG------------------GTTGATGAA------------------------------ 4427

Qy        326 AspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHis 345
                                 |||:::   ||||||            :::      |||
Db       4428 ------------------AAGCGCCGTAAAACCAAGCGTGCTGAAAATCGCGCTCGCGGT 4469

Qy        346 GlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHis 365
                           ||||||   ||||||                              |||
Db       4470 CGTGCAGAGCGTAAAGCTAAACGCCAAACA------------------------TTTTAA 4505

Qy        366 PheAlaLeu------------GlnProAsnPheLeuGlyMetCysThrAsnTyrLysGlu 381
                 |||               |||   ||||||   |||   |||||||||
Db       4506 TAGAAATTGTTGGAGTATTGAGCTCCAACTTTTTTATTTATGAGAACGAACTATCTAAAC 4565

Qy        382 ArgLeuCysTyrIleAsnAsnSer-ValSerAsnLysProAlaIleIleLeuSerSerLe 401
              |||         :::      ::: ::::::   |||      |||::::::
Db       4566 CGAAACACTACATTAAAGACTGCAAATTGCGATTAAAAATGGTATAATGATA-------- 4617

Qy        401 uValGlyAsnLeuValAspGlnSerLysGln---GlyIleValPheAsnGluAlaSerTr 420
                          |||         |||      ||||||   :::   |||   |||::
Db       4618 -------AAGTTATATAGTCCCGATAAGATGGTAGGTATTTATTACGAAGAGTTTTCCTA 4670

Qy        420 p---------AlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAsp-- 436
              :                |||   :::   |||   |||   :::   :::   :::
Db       4671 TCAGTACTTTGTAACTCTATAACAATATTTTTTAAGGGGGGACATTTTTATGTCAGAGCG 4730

Qy        437 -ProMetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTy 456
                 ::::::   |||:::|||   :::         |||   |||      |||   |||
Db       4731 TAAATTATTCACGTCTGAATCT---GTATCTGAGGGGCATCCGGATAAGATTGCAGACCA 4787

Qy        456 rLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAl 476
               :::      :::|||            :::   |||:::   ||||||   |||   ||
Db       4788 AATTTCAGATGCGATTTTGGATGCTATTTTAGCAAAGGATCCAGAGGCG---CACGTTGC 4844

Qy        476 aMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSe 496
              |   :::   |||      |||               |||   :::         :::
Db       4845 TGCTGAAACAGCTGTATATACTGGTTCTGTCCACGTTTTTGGTGAAATTTCTACAAATGC 4904

Qy        496 rTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheTh 516
              :|||            :::   :::   |||
Db       4905 CTATGTGGATATTAACCGTGTGGTTCGTGAT----------------------------- 4935

Qy        516 rThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGly 530
               |||:::         ||||||         |||||||||
Db       4936 -ACCATTGCA---GAGATTGGTTATACCAATACAGAATATGGA 4974
```

RESULT 6
US-08-987-439-2
; Sequence 2, Application US/08987439
; Patent No. 6110457
; GENERAL INFORMATION:
; APPLICANT: Belshe, Robert B.
; APPLICANT: Ray, Ranjit
; TITLE OF INVENTION: LIVE ATTENUATED VACCINES BASED ON cp45 HPIV-3 STRAIN
; TITLE OF INVENTION: AND METHOD TO ENSURE ATTENUATION IN SUCH VACCINES
; FILE REFERENCE: SLU 4481
; CURRENT APPLICATION NUMBER: US/08/987,439
; CURRENT FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 6702
;   TYPE: DNA
;   ORGANISM: Human parainfluenza virus 3
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(6702)
US-08-987-439-2

Alignment Scores:

Pred. No.:	0.0375	Length:	6702
Score:	112.50	Matches:	128
Percent Similarity:	35.57%	Conservative:	100
Best Local Similarity:	19.97%	Mismatches:	214
Query Match:	3.74%	Indels:	199
DB:	3	Gaps:	33

US-09-913-878A-2_COPY_710_1282 (1-573) x US-08-987-439-2 (1-6702)

```
Qy         40 AspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrp 59
              |||:::||||||         |||            :::   |||         ::::::|||
Db        127 GATATGGATGAC------GACTCAATACTAGTTATCACT---------AGACAGAAAATA 171

Qy         60 GluCysAspPheValAspLysHisGlnArgThrLeu------------------------ 71
              :::   :::   :::||||||      ||||||:::::::
Db        172 AAACTTAATAAATTGGATAAAAGACAACGATCTATTAGAAGATTAAAATTAATATTAACT 231

Qy         72 ------------------------------------------GluValArgSerVal 76
                                                        |||:::
Db        232 GAAAAAGTGAATGACTTAGGAAAATACACATTTATCAGATATCCAGAAATGTCAAAAGAA 291

Qy         77 AlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArg 96
                    :::|||            |||:::|||      ::::::         :::|||
Db        292 ATGTTCAAATTATATATACCTGGTATTAACAGTAAAGTGACTGAATTATTACTTAAAGCA 351

Qy         97 AlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGln 116
                 |||         :::|||               :::   |||         ||||||   |||
Db        352 GATAGAACATATAGTCAAATGACTGATGGATTAAGAGATCTATGGATTAATGTGCTATCA 411

Qy        117 ArgGlnPheSerGlu---------GlnLysHisAlaLeuAsnArgProValGlu------ 131
              :::         |||:::               :::      ||||||      :::
Db        412 AAATTAGCCTCAAAAAATGATGGAAGCAATTATGATCTTAATGAAGAAATTAATAATATA 471

Qy        132 ------------------------------------------PheArgGlnTrpVal 136
                                                        |||:::   |||
Db        472 TCGAAAGTTCACACAACCTATAAATCAGATAAATGGTATAATCCATTCAAAACATGGTTT 531

Qy        137 TyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAla 156
                       |||      |||         :::   :::         :::   |||
Db        532 ACTATCAAGTATGATATGAGAAGATTACAAAAAGCTCGAAATGAGATCACTTTTAATGTT 591

Qy        157 GlyLeuPro---AspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAsp 175
              |||            :::      |||:::         |||||||||:::
Db        592 GGGAAGGATTATAACTTGTTAGAAGACCAGAAGAATTTCTTATTG---------ATACAT 642

Qy        176 ProLys------------------LysGlnLysTyrLeu---GlnAspIleAlaTrpAsp 188
              |||:::                  ||||||   |||         |||         :::
Db        643 CCAGAATTGGTTTTGATATTAGATAAACAAAACTATAATGGTTATCTAATTACTCCTGAA 702
```

```
Qy        189 LeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgVal--------- 205
              |||         |||||   ::::::  :::   |||||
Db        703 TTAGTATTGATGTATTGTGACGTAGTCGAAGGCCGATGGAATATAAGTGCATGTGCTAAG 762

Qy        206 ------GlyArgSerAlaTyrIleTyrMetIleAla---AspPheTrpGlyValLeuGlu 222
                    :::         :::|||       :::  |||  |||::::::
Db        763 TTAGATCCAAAATTACAATCTATGTATCAGAAAGGTAATAACCTGTGGGAAGTGATAGAT 822

Qy        223 GluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheThr 242
              :::          :::|||                       |||::::::|||
Db        823 AAATTGTTTCCAATTATGGGA-----------------------GAAAAGACATTTGAT 858

Qy        243 LeuLeuSerAspCysAspValLeu---ValAlaArgSerProAlaHisPheProSerAsp 261
              ::::::|||     :::   |||  ::::::             |||   |||
Db        859 GTGATATCGTTATTAGAACCACTTGCATTATCCTTAATTCAAACTCATGATCCT------ 912

Qy        262 IleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal---Ile 280
              ::::::::::::|||    |||          |||   |||  ::::::    :::
Db        913 GTTAAACAACTAAGAGGAGCTTTTTTAAAT------CATGTGTTATCCGAGATGGAATTA 966

Qy        281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 300
              ||||||  ::::::     :::          |||:::          ||||||
Db        967 ATATTTGAATCTAGAGAATCGATT---------AAGGAATTTCTGAGTGTAGATTAC--- 1014

Qy        301 GlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGlu 320
                                           :::|||:::|||   |||     :::
Db       1015 ------------------------ATTGATAAAATTTAGATATATTTAATAAGTCTACA 1050

Qy        321 MetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPhe------ 338
              :::       ::::::      :::   :::                |||||
Db       1051 ATAGATGAAATAGCAGAGATTTTCTCTTTTTTAGA------------ACATTTGGGCAT 1098

Qy        339 LysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLys---------------Glu 353
                    |||   ||||||                    |||                 :::
Db       1099 CCTCCATTAGAAGCTAGTATTGCAGCAGAAAAGGTTAGAAAATATATGTATATTGGAAAA 1158

Qy        354 GlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeu 373
              |||      :::|||   |||   |||
Db       1159 CAATTAAAATTTGACACTATTAATAAATGTCATGCTATC--------------------- 1197

Qy        374 GlyMetCysThr------------AsnTyrLysGluArgLeuCys--------------- 384
                    |||||              |||:::|||||
Db       1198 ---TTCTGTACAATAATAATTAACGGATATAGAGAGAGGCATGGTGGACAGTGGCCTCCT 1254

Qy        385 --------------------------TyrIleAsnAsnSerValSerAsnLysProAla 395
                                        :::|||   |||:::      :::    |||
Db       1255 GTGACATTACCTGATCATGCACACGAATTCATCATAAATGCTTACGGTTCAAACTCTGCG 1314

Qy        396 IleIleLeuSerSerLeuValGlyAsnLeuValAsp------GlnSerLysGlnGlyIle 413
              |||                   |||   ||||||     ||||||     ||||||
Db       1315 ATATCATATGAA-----------AATGCTGTTGATTATTACCAGAGCTTTATAGGAATA 1362

Qy        414 ValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeu------------------ 427
                 ||||||:::               ||||||     :::|||
Db       1363 AAATTCAATAAATTCATAGAGCCTCAGTTAGATGAGGATTTGACAATTTATATGAAAGAT 1422

Qy        428 ------LeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAspSerTrpLeu 445
                    |||         |||      |||   :::|||   :::   :::   |||
Db       1423 AAAGCATTATCTCCAAAAAAATCAAATTGGGACACAGTTTATCCTGCATCTAATTTACTG 1482

Qy        446 GlyArgGly--------------------------------------------------- 448
                 |||
Db       1483 TACCGTACTAACGCATCCAACGAATCACGAAGATTAGTTGAAGTATTTATAGCAGATAGT 1542

Qy        449 ------GluProThrHisIleIleAspTyrLeu-------------------------- 457
                    :::|||      |||:::||||||:::
Db       1543 AAATTTGATCCTCATCAGATATTGGATTATGTAGAATCTGGGGACTGGTTAGATGATCCA 1602

Qy        458 LysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMet 477
              :::|||:::|||:::                :::||||||::::::
Db       1603 GAATTTAATATTTCTTATAGTCTTAAAGAAAAAGAGATCAAA------------------ 1644

Qy        478 LysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyr 497
                                 ::::::|||
Db       1645 ------------------CAGGAAGGT-------------------------------- 1653

Qy        498 TyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPhe---Thr 516
                 |||   :::::::::   |||   |||::::::    :::|||      |||
Db       1654 ---AGACTCTTTGCAAAAATGACATACAAAATGAGAGCTACACAAGTTTTATCAGAGACA 1710

Qy        517 ThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLys 536
                 |||   |||   ||||||:::      :::|||   |||    :::
Db       1711 CTACTTGCAAATAACATAGGAAAATTCTTTCAAGAAAATGGGATGGTGAAGGGAGAGATT 1770

Qy        537 GluMetArgAspSerLysAspProTyrProVal-----------ArgValAsnGlnVal 552
              |||:::                           :::           |||   |||:::|||
Db       1771 GAATTACTTAAGAGATTAACAACCATATCAATATCAGGAGTTCCACGGTATAATGAAGTG 1830

Qy        553 Tyr 553
              |||
Db       1831 TAC 1833
```

```
RESULT 7
US-08-750-717-1
; Sequence 1, Application US/08750717
; Patent No. 6180109
;  GENERAL INFORMATION:
;    APPLICANT:  MOORMANN, Robertus J. M.
;    APPLICANT:  VAN RIJN, Petrus A.
;    TITLE OF INVENTION:  Nucleotide Sequences of Pestivirus
;    TITLE OF INVENTION:  Strains, Polypeptides Encoded by These Sequences and Use
;    TITLE OF INVENTION:  Thereof for Diagnosis and Prevention of Pestivirus
;    TITLE OF INVENTION:  Infections
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  YOUNG & THOMPSON
;      STREET:  745 South 23rd Street
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/750,717
;      FILING DATE:  24-DEC-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 94201743.5
;      FILING DATE:  17-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/NL95/00214
;      FILING DATE:  16-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PATCH, Andrew J.
;      REGISTRATION NUMBER:  32,925
;      REFERENCE/DOCKET NUMBER:  BO 39123
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-521-2297
;      TELEFAX:  703-685-0573
;      TELEX:  248425 EMBON
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12311 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  374..12067
US-08-750-717-1
```

```
Alignment Scores:
Pred. No.:                0.153        Length:        12311
Score:                    111.00       Matches:       102
Percent Similarity:       35.48%       Conservative:  80
Best Local Similarity:    19.88%       Mismatches:    187
Query Match:              3.69%        Indels:        144
DB:                       3            Gaps:          23

US-09-913-878A-2_COPY_710_1282 (1-573) x US-08-750-717-1 (1-12311)

Qy        67 HisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsn 86
             |||:::      :::      ::::::::::::|||   :::   :::   |||:::
Db     10301 CACAGACACAATGTGTATAACAAGACAATAAGCTCAGTAATGACAGCTACTGGTATCAGG 10360

Qy        87 LeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAla 106
             |||:::   |||||||||:::      :::      |||            ||||||
Db     10361 TTGGAGAAGTTGCCCGTGGTTAGGGCCCAGACA---GACCCAACCAACTTCCACCAAGCA 10417

Qy       107 IleGlyAspArgLeuIleAsnAsp--------------------LeuGlnArgGlnPhe 119
             |||   |||::::::      :::                    |||   ::::::
Db     10418 ATAAGGGATAAGATAGACAAGGAAGAGAACCTACAAACCCCGGGTTTACATAAGAAATTA 10477

Qy       120 SerGluGlnLysHisAlaLeuAsnArgPro-----------------------ValGlu 131
                |||   :::||||||   ||||||                          ||||||
Db     10478 ATGGAAGTTTTCAACGCATTGAAACGACCCGAGTTAGAGTCCTCCTACGACGCCGTGGAA 10537

Qy       132 Phe-------------------------------------------------ArgGln 134
             :::                                                 |||:::
Db     10538 TGGGAGGAACTGGAGAGAGGAATAAACAGGAAGGGTGCTGCTGGTTTTTCGAACGCAAA 10597

Qy       135 TrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHis--------------- 149
                :::   |||         |||         :::|||
Db     10598 AATATAGGGGAAATATTGGATTCAGAGAAAAATAAAGTCGAAGAGATTATTGACAATCTG 10657

Qy       150 ------GlyArg------ValProPheLeuAlaGlyLeuProAspSerGlnGluGluThr 165
                   ||||||                   :::   :::|||   ::::::::
Db     10658 AAAAAAGGTAGAAACATTAAATATTATGAAACCGCGATCCCAAAGAATGAGAAGAGGGAC 10717

Qy       166 LeuAsnPheLeuMetAsnSerGly------PheAspProLysLysGlnLysTyrLeuGln 183
             :::|||      :::|||            |||   ||||||   :::   :::|||
Db     10718 GTCAACGATGACTGGACCGCCGGTGATTTCGTGGACGAGAAGAAACCTAGAGTCATACAA 10777

Qy       184 -----------AspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSer 199
                                   :::   |||               :::|||
Db     10778 TACCCTGAAGCAAAAACAAGACTGGCCATCACCAAGGTGATGTATAAGTGGGTGAAGCAG 10837

Qy       200 LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219
             |||         :::   |||                     |||
Db     10838 AAGCCAGTAGTTATACCCGGGTATGAAGGGAAGACACCTCTATTCCAAATTTTTGACAAA 10897

Qy       220 ValLeuGluGlu---------------AsnGluValHisValGlyPheSerSerLysPhe 234
             |||   :::|||               |||   |||   |||   |||   :::|||
Db     10898 GTAAAGAAGGAATGGATCAATTTCAAAATCCAGTGGCAGTGAGCTTCGACACTAAGGCG 10957

Qy       235 ArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSer 254
                |||                |||   ::::::   |||:::   :::   |||
Db     10958 TGGGAC---------------ACCCAGGTAACCACAAAAGATTTGGAGCTGATAAGG--- 10999

Qy       255 ProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHis 274
                                ||||||||:::         ||||||   :::   |||
Db     11000 ------------------GACATACAAAAGTAT------TATTTCAAGAAGAAATGGCAC 11035

Qy       275 Ser-----------------------LeuLysAspValIleIlePheSerThrLysGly 286
                                        :::   :::|||   :::   |||      |||
Db     11036 AAATTTATTGACACCCTGACCACGCATATGTCAGAAGTACCCGTGATCAGTGCTGATGGG 11095

Qy       287 AspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAla------ 304
             :::|||   :::   |||         |||                     |||
Db     11096 GAAGTATACATAAGGAAAGGGCAAAGAGGCAGTGGACAACCTGACACAAGTGCGGGCAAC 11155

Qy       304 ------------------------------------------------------------ 304

Db     11156 AGCATGCTAAATGTCTTAACAATGGTTTACGCCTTCTGCGAGGCCACAGGAGTACCCTAC 11215

Qy       305 -------------------------TrpValCysTrpAspProGluIleValAspGly 315
                                      ||||||   |||            ||||||
Db     11216 AAGAGCTTTGACAGGGTGGCAAAAATTCATGTGTGCGGGGAT-----------GATGGC 11263

Qy       316 PheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLys 335
             |||:::      |||      |||                           :::|||
Db     11264 TTCCTGATCACAGAAAGAGCTCTC--------------------------GGTGAGAAA 11296

Qy       336 ThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThr 355
                   |||   :::         :::         |||      |||         |||
Db     11297 TTTGCAAGTAAGGGAGTCCAGATCCTTTATGAAGCTGGGAAGCCCCAGAAG---ATCACT 11353

Qy       356 ThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMet 375
                   |||   ::::::   ::::::   |||                :::   :::
Db     11354 GAAGGGGACAAAATGAAAGTGGCCTACCAATTT-----------GATGATATTGAGTTT 11401

Qy       376 CysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysPro--- 394
             |||::::::         ::::::   :::   :::|||:::   |||      |||
Db     11402 TGCTCCCATACACCAATACAAGTAAGATGGTCAGATAACACTTCTAGTTACATGCCGGGG 11461

Qy       395 ---------AlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGln 411
                      ||||||:::   :::         |||                  ::::::
Db     11462 AGAAATACAACCACAATCCTGGCAAAGATGGCCACGAGGTTAGATTCCAGCGGTGAAAGG 11521

Qy       412 Gly---IleValPheAsnGluAla-------------------------SerTrpAla 421
             |||   |||   :::   :::|||                          ||||||
Db     11522 GGTACCATAGCATATGAGAAAGCAGTAGCATTTAGCTTCCTGCTGATGTACTCCTGGAAC 11581

Qy       422 GlnLeuArgArgGlu------LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyr 439
                |||   |||            ||||||         :::      |||
Db     11582 CCACTAATTAGAAGGATCTGCTTACTGGTGCTATCAACTGAACTGCAAGTGAAACCAGGG 11641

Qy       440 LysSerAspSerTrpLeuGlyArgGlyGluPro-----------ThrHisIleIleAsp 455
             ||||||   ::::::         |||:::|||                 :::|||
Db     11642 AAGTCAACTACTTACTATTATGAAGGAGACCCGATATCTGCCTACAAGGAAGTCATCGGC 11701

Qy       456 TyrLeuLysPheSerIleAlaArgProAlaIleAspLys 468
             :::         |||   :::   |||   :::   :::|||
Db     11702 CACAACCTTTTTGATCTTAAGAGAACAAGCTTTGAGAAG 11740

RESULT 8
US-08-804-196-1
; Sequence 1, Application US/08804196
; Patent No. 5874256
;  GENERAL INFORMATION:
;    APPLICANT:  Bertina, Rogier
;    APPLICANT:  Reitsma, Pieter
;    TITLE OF INVENTION:  A method for diagnosing an increased
;    TITLE OF INVENTION:  risk for thrombosis or a genetic defect causing
;    TITLE OF INVENTION:  thrombosis and kit for use with the same.
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Akzo No. 5874256el Patent Department
;      STREET:  1300 Piccard Drive, Suite 206
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,196
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/454,353
;      FILING DATE:  06-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gormley, Mary E.
```

```
;     REGISTRATION NUMBER:  34,409
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (301) 948-7400
;     TELEFAX:  (301) 948-9751
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  6909 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  both
;     TOPOLOGY:  unknown
;   MOLECULE TYPE:  cDNA
US-08-804-196-1

Alignment Scores:
Pred. No.:                0.0681      Length:         6909
Score:                    110.50      Matches:        109
Percent Similarity:       37.12%      Conservative:   74
Best Local Similarity:    22.11%      Mismatches:     199
Query Match:              3.67%       Indels:         112
DB:                       2           Gaps:           28

US-09-913-878A-2_COPY_710_1282 (1-573) x US-08-804-196-1 (1-6909)

Qy        28 GlyArgPhe------GlySerAlaLysGlyMetTrpValIle------AspValAsp--- 42
              |||||:::       |||   ||||||   :::   |||       |||  |||
Db      2923 GGGAGATGGCATTTGGCTTCTGAGAAAGGTAGCTATGAAATAATCCAAGATACTGATGAA 2982

Qy        43 AspThrGlyAspGluAspTrp---IleGluThrTyrProSerGlnArgLysTrp------ 59
              |||||          :::|||   |||                :::  |||  |||
Db      2983 GACACAGCTGTTAACAATTGGCTGATCAGCCCCCAGAATGCCTCACGTGCTTGGGGAGAA 3042

Qy        60 GluCysAspPheValAspLys------------------HisGlnArgThrLeuGluVal 73
                               :::|||                  |||   :::      |||
Db      3043 AGCACCCCTCTTGCCAACAAGCCTGGAAAGCAGAGTGGCCACCCAAAGTTTCCTAGAGTT 3102

Qy        74 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 93
              |||                                         :::  |||  |||
Db      3103 AGACAT------------------------------------AAATCTCTACAAGTAAGA 3126

Qy        94 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 113
              :::|||     :::  ::::::|||      |||             ||||||
Db      3127 CAGGATGGAGGAAAGAGTAGACTGAAGAAAAGCCAGTTT-----------CTCATTAAG 3174

Qy       114 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu------ 131
                      |||:::       ::::::||||||      :::   |||:::
Db      3175 ------ACACGAAAAAAGAAAAAAGAGAAGCACACACACCATGCTCCTTTATCTCCGAGG 3228

Qy       132 ------------PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgVal 147
                         |||      |||   ::::::|||  |||      |||:::
Db      3229 ACCTTTCACCCTCTAAGAAGTGAAGCCTACAACACATTTTCAGAAAGA------AGACTT 3282

Qy       148 SerHisGlyArgValProPheLeuAla---------GlyLeuProAspSerGlnGluGlu 164
                 |||   |||       :::           ||||||            :::
Db      3283 AAGCATTCGTTGGTGCTTCATAAATCCAATGAAACATCTCTTCCCACAGACCTCAATCAG 3342

Qy       165 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAsp 184
              ||||||       |||:::    |||:::                :::  |||:::
Db      3343 ACATTGCCC---TCTATGGATTTTGGCTGGATAGCCTCACTTCCTGACCATAATCAGAAT 3399

Qy       185 IleAlaTrpAspLeuGlnLysArgLysCys--------------AspThrLeuLysSer 199
                 :::  |||       :::     |||                 |||:::
Db      3400 TCCTCAAATGACACTGGTCAGGCAAGCTGTCCTCCAGGTCTTTATCAGACAGTGCCCCCA 3459

Qy       200 LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219
              :::   :::   :::                                       :::
Db      3460 GAGGAACACTATCAAACA-----------------------------------TTCCCC 3483

Qy       220 ValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGlu 239
              :::   :::   :::::::::|||                  |||
Db      3484 ATTCAAGACCCTGATCAAATGCACTCTACTTCAGACCCAGTCACAGATCCTCTTCTCCA 3543

Qy       240 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 259
                 :::   :::   :::  |||               ||||||        ||||||
Db      3544 GAGCTCAGTGAAATGCTTGAGTATGAC-----------CGAAGTCACAAGTCCTTCCCC 3591

Qy       260 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 279
              :::||||||   ::::::                :::         |||   :::
Db      3592 ACAGATATAAGTCAAATGTCCCCTTCCTCAGAA---------CATGAAGTCTGGCAGACA 3642

Qy       280 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyr 299
              :::|||                        |||   |||:::   :::||||||       :::
Db      3643 GTCATCTCTCCAGACCTCAGCCAGGTGACCCTCTCTCCAGAACTCAGCCAGACAAACCTC 3702

Qy       300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsn--- 318
                    |||::::::              ||||||::::::              ::::::
Db      3703 TCTCCAGACCTCAGCCACACGACTCTCTCTCCAGAACTCATTCAGAGAAACCTTTCCCCA 3762

Qy       319 ------AlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThr 336
                    :::||||||:::   ||||||||||||                        |||
Db      3763 GCCCTCGGTCAGATGCCCATTTCTCCAGACCTCAGCCAT-----------------ACA 3804

Qy       337 ThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThr 356
              |||             :::   ||||||   |||   |||            ||||||
Db      3805 ACCCTTTCTCCAGACCTC---AGCCATACAACCCTTTCTTTAGACCTCAGCCAGACAAAC 3861

Qy       357 Tyr------AspMetIleGlnLysSerPhe---------------------------- 364
                       ::::::  |||   :::
Db      3862 CTCTCTCCAGAACTCAGTCAGACAAACCTTTCCCCAGCCCTCGGTCAGATGCCCCTTTCT 3921

Qy       365 ------------HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
                          |||     |||        :::|||                ||||||
Db      3922 CCAGACCTCAGCCATACAACCCTTTCTCTAGACTTC------AGCCAGACAAACCTCTCT 3975

Qy       381 GluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIle--------- 397
                 |||  ::::::       |||                 |||   :::
Db      3976 CCAGAACTCAGCCATATGACTCTCTCT------------CCAGAACTCAGTCAGACAAAC 4023

Qy       398 LeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGlu 417
              ||||||       :::|||     :::           |||:::           :::|||
Db      4024 CTTTCCCCAGCCCTTGGTCAGATGCC-CATTTCTCCAGACCTCAG---CCATACAACCCT 4079

Qy       418 AlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspPro 437
                 |||    |||     :::   |||                :::   |||   |||
Db      4080 TTCTCTAGACTTCAGCCAGACAAACCTCTCTCCAGAACTCAGTCAAACAAACCTTTCCCC 4139

Qy       438 MetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeu 457
                    :::||||||:::         |||   :::|||                  |||
Db      4140 AGCCCTCGGTCAGATGCCCCTTTCTCCAGACCCCAGCCA---------------TACAAC 4184

Qy       458 LysPheSerIleAlaArgProAlaIleAspLysGluLeu 470
                 ||||||         ||||||         ||||||   |||
Db      4185 CCTTTCTCT------AGACCTCAGCCAGACAAACCTCTC 4217

RESULT 9
US-08-658-340-1
; Sequence 1, Application US/08658340
; Patent No. 5910576
; Patent No. 5910576 5861489
;  GENERAL INFORMATION:
;    APPLICANT:  Bertina, Rogier
;    APPLICANT:  Reitsma, Pieter
;    TITLE OF INVENTION:  A method for diagnosing an increased
;    TITLE OF INVENTION:  risk for thrombosis or a genetic defect causing
;    TITLE OF INVENTION:  thrombosis and kit for use with the same.
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Akzo No. 5910576 5861489el Patent Department
;      STREET:  1300 Piccard Drive, Suite 206
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/658,340
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/454,353
;      FILING DATE:  06-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gormley, Mary E.
;      REGISTRATION NUMBER:  34,409
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 948-7400
;      TELEFAX:  (301) 948-9751
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6909 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-658-340-1

Alignment Scores:
Pred. No.:               0.0681      Length:          6909
Score:                   110.50      Matches:         109
Percent Similarity:      37.12%      Conservative:    74
Best Local Similarity:   22.11%      Mismatches:      199
Query Match:             3.67%       Indels:          112
DB:                      2           Gaps:            28

US-09-913-878A-2_COPY_710_1282 (1-573) x US-08-658-340-1 (1-6909)

Qy       28 GlyArgPhe------GlySerAlaLysGlyMetTrpValIle------AspValAsp--- 42
Db     2923 GGGAGATGGCATTTGGCTTCTGAGAAAGGTAGCTATGAAATAATCCAAGATACTGATGAA 2982
Qy       43 AspThrGlyAspGluAspTrp---IleGluThrTyrProSerGlnArgLysTrp------ 59
Db     2983 GACACAGCTGTTAACAATTGGCTGATCAGCCCCCAGAATGCCTCACGTGCTTGGGGAGAA 3042
Qy       60 GluCysAspPheValAspLys-----------------HisGlnArgThrLeuGluVal 73
Db     3043 AGCACCCCTCTTGCCAACAAGCCTGGAAAGCAGAGTGGCCACCCAAAGTTTCCTAGAGTT 3102
Qy       74 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 93
Db     3103 AGACAT-----------------------------------AAATCTCTACAAGTAAGA 3126
Qy       94 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 113
Db     3127 CAGGATGGAGGAAAGAGTAGACTGAAGAAAAGCCAGTTT------------CTCATTAAG 3174
Qy      114 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu------ 131
Db     3175 ------ACACGAAAAAAGAAAAAAGAGAAGCACACACACCATGCTCCTTTATCTCCGAGG 3228
Qy      132 ------------PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgVal 147
Db     3229 ACCTTTCACCCTCTAAGAAGTGAAGCCTACAACACATTTTCAGAAAGA------AGACTT 3282
Qy      148 SerHisGlyArgValProPheLeuAla---------GlyLeuProAspSerGlnGluGlu 164
Db     3283 AAGCATTCGTTGGTGCTTCATAAATCCAATGAAACATCTCTTCCCACAGACCTCAATCAG 3342
Qy      165 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAsp 184
Db     3343 ACATTGCCC---TCTATGGATTTTGGCTGGATAGCCTCACTTCCTGACCATAATCAGAAT 3399
Qy      185 IleAlaTrpAspLeuGlnLysArgLysCys---------------AspThrLeuLysSer 199
Db     3400 TCCTCAAATGACACTGGTCAGGCAAGCTGTCCTCCAGGTCTTTATCAGACAGTGCCCCCA 3459
Qy      200 LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219
Db     3460 GAGGAACACTATCAAACA-----------------------------------TTCCCC 3483
Qy      220 ValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGlu 239
Db     3484 ATTCAAGACCCTGATCAAATGCACTCTACTTCAGACCCCAGTCACAGATCCTCTTCTCCA 3543
Qy      240 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 259
Db     3544 GAGCTCAGTGAAATGCTTGAGTATGAC------------CGAAGTCACAAGTCCTTCCCC 3591
Qy      260 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 279
Db     3592 ACAGATATAAGTCAAATGTCCCCTTCCTCAGAA---------CATGAAGTCTGGCAGACA 3642
Qy      280 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyr 299
Db     3643 GTCATCTCTCCAGACCTCAGCCAGGTGACCCTCTCTCCAGAACTCAGCCAGACAAACCTC 3702
Qy      300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsn--- 318
Db     3703 TCTCCAGACCTCAGCCACACGACTCTCTCTCCAGAACTCATTCAGAGAAACCTTTCCCCA 3762
Qy      319 ------AlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThr 336
Db     3763 GCCCTCGGTCAGATGCCCATTTCTCCAGACCTCAGCCAT------------------ACA 3804
Qy      337 ThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThr 356
Db     3805 ACCCTTTCTCCAGACCTC---AGCCATACAACCCTTTCTTTAGACCTCAGCCAGACAAAC 3861
Qy      357 Tyr------AspMetIleGlnLysSerPhe----------------------------- 364
Db     3862 CTCTCTCCAGAACTCAGTCAGACAAACCTTTCCCCAGCCCTCGGTCAGATGCCCCTTTCT 3921
Qy      365 ------------HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
Db     3922 CCAGACCTCAGCCATACAACCCTTTCTCTAGACTTC------AGCCAGACAAACCTCTCT 3975
Qy      381 GluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIle--------- 397
Db     3976 CCAGAACTCAGCCATATGACTCTCTCT------------CCAGAACTCAGTCAGACAAAC 4023
Qy      398 LeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGlu 417
Db     4024 CTTTCCCCAGCCCTTGGTCAGATGCC-CATTTCTCCAGACCTCAG---CCATACAACCCT 4079
Qy      418 AlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspPro 437
Db     4080 TTCTCTAGACTTCAGCCAGACAAACCTCTCTCCAGAACTCAGTCAAACAAACCTTTCCCC 4139
Qy      438 MetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeu 457
Db     4140 AGCCCTCGGTCAGATGCCCCTTTCTCCAGACCCCAGCCA---------------TACAAC 4184
Qy      458 LysPheSerIleAlaArgProAlaIleAspLysGluLeu 470
Db     4185 CCTTTCTCT------AGACCTCAGCCAGACAAACCTCTC 4217

RESULT 10
US-08-746-111-26
; Sequence 26, Application US/08746111
; Patent No. 6066778
;  GENERAL INFORMATION:
;    APPLICANT:  Ginsburg, David
;    APPLICANT:  Cui, Jisong
;    TITLE OF INVENTION:  Compositions And Methods For Screening
;    TITLE OF INVENTION:  Compounds For Anticoagulant Activity
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
```

```
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/746,111
;      FILING DATE:  06-NOV-1996
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ingolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  UM-02536
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6909 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-746-111-26

Alignment Scores:
Pred. No.:              0.0681      Length:         6909
Score:                  110.50      Matches:        109
Percent Similarity:     37.12%      Conservative:   74
Best Local Similarity:  22.11%      Mismatches:     199
Query Match:            3.67%       Indels:         112
DB:                     3           Gaps:           28

US-09-913-878A-2_COPY_710_1282 (1-573) x US-08-746-111-26 (1-6909)

Qy         28 GlyArgPhe------GlySerAlaLysGlyMetTrpValIle------AspValAsp--- 42
              ||||||:::          |||  ||||||   :::  |||      |||   |||
Db       2923 GGGAGATGGCATTTGGCTTCTGAGAAAGGTAGCTATGAAATAATCCAAGATACTGATGAA 2982

Qy         43 AspThrGlyAspGluAspTrp---IleGluThrTyrProSerGlnArgLysTrp------ 59
              ||||||         :::|||   |||             :::   |||   |||
Db       2983 GACACAGCTGTTAACAATTGGCTGATCAGCCCCCAGAATGCCTCACGTGCTTGGGGAGAA 3042

Qy         60 GluCysAspPheValAspLys-----------------HisGlnArgThrLeuGluVal 73
                                 :::|||                 |||   :::       |||
Db       3043 AGCACCCCTCTTGCCAACAAGCCTGGAAAGCAGAGTGGCCACCCAAAGTTTCCTAGAGTT 3102

Qy         74 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 93
              |||                                            :::   |||   |||
Db       3103 AGACAT------------------------------------AAATCTCTACAAGTAAGA 3126

Qy         94 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 113
              :::|||      :::   ::::::|||         |||               ||||||
Db       3127 CAGGATGGAGGAAAGAGTAGACTGAAGAAAAGCCAGTTT------------CTCATTAAG 3174

Qy        114 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu------ 131
                       |||:::      ::::::||||||         :::   |||:::
Db       3175 ------ACACGAAAAAAGAAAAAAGAGAAGCACACACACCATGCTCCTTTATCTCCGAGG 3228

Qy        132 ------------PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgVal 147
                                |||      |||   ::::::|||   |||      |||:::
Db       3229 ACCTTTCACCCTCTAAGAAGTGAAGCCTACAACACATTTTCAGAAAGA------AGACTT 3282

Qy        148 SerHisGlyArgValProPheLeuAla---------GlyLeuProAspSerGlnGluGlu 164
                 |||      |||         :::            ||||||            :::
Db       3283 AAGCATTCGTTGGTGCTTCATAAATCCAATGAAACATCTCTTCCCACAGACCTCAATCAG 3342

Qy        165 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAsp 184
              ||||||            |||:::   |||:::                  :::   |||:::
Db       3343 ACATTGCCC---TCTATGGATTTTGGCTGGATAGCCTCACTTCCTGACCATAATCAGAAT 3399

Qy        185 IleAlaTrpAspLeuGlnLysArgLysCys---------------AspThrLeuLysSer 199
                 :::   |||      :::      |||                     |||:::
Db       3400 TCCTCAAATGACACTGGTCAGGCAAGCTGTCCTCCAGGTCTTTATCAGACAGTGCCCCCA 3459

Qy        200 LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219
              :::   :::   :::                                          :::
Db       3460 GAGGAACACTATCAAACA------------------------------------TTCCCC 3483

Qy        220 ValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGlu 239
              :::   :::   :::::::::|||                     |||
Db       3484 ATTCAAGACCCTGATCAAATGCACTCTACTTCAGACCCCAGTCACAGATCCTCTTCTCCA 3543

Qy        240 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 259
                 :::   :::   :::   |||                  ||||||         ||||||
Db       3544 GAGCTCAGTGAAATGCTTGAGTATGAC------------CGAAGTCACAAGTCCTTCCCC 3591

Qy        260 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 279
              :::||||||   ::::::                  :::            |||   :::
Db       3592 ACAGATATAAGTCAAATGTCCCCTTCCTCAGAA---------CATGAAGTCTGGCAGACA 3642

Qy        280 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyr 299
              :::|||               |||   |||:::   :::||||||            :::
Db       3643 GTCATCTCTCCAGACCTCAGCCAGGTGACCCTCTCTCCAGAACTCAGCCAGACAAACCTC 3702

Qy        300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsn--- 318
                    |||::::::                  ||||||::::::            ::::::
Db       3703 TCTCCAGACCTCAGCCACACGACTCTCTCTCCAGAACTCATTCAGAGAAACCTTTCCCCA 3762

Qy        319 ------AlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThr 336
                       :::||||||:::   ||||||||||||                     |||
Db       3763 GCCCTCGGTCAGATGCCCATTTCTCCAGACCTCAGCCAT------------------ACA 3804

Qy        337 ThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThr 356
              |||            :::   ||||||   |||   |||            ||||||
Db       3805 ACCCTTTCTCCAGACCTC---AGCCATACAACCCTTTCTTTAGACCTCAGCCAGACAAAC 3861

Qy        357 Tyr------AspMetIleGlnLysSerPhe----------------------------- 364
                       ::::::   |||   :::
Db       3862 CTCTCTCCAGAACTCAGTCAGACAAACCTTTCCCCAGCCCTCGGTCAGATGCCCCTTTCT 3921

Qy        365 ------------HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
                          |||   |||   :::|||               ||||||
Db       3922 CCAGACCTCAGCCATACAACCCTTTCTCTAGACTTC------AGCCAGACAAACCTCTCT 3975

Qy        381 GluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIle--------- 397
                 |||   ::::::      |||            |||   :::
Db       3976 CCAGAACTCAGCCATATGACTCTCTCT------------CCAGAACTCAGTCAGACAAAC 4023

Qy        398 LeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGlu 417
              ||||||      :::|||         :::   |||:::            :::|||
Db       4024 CTTTCCCCAGCCCTTGGTCAGATGCC-CATTTCTCCAGACCTCAG---CCATACAACCCT 4079

Qy        418 AlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspPro 437
                 |||   |||      :::   |||               :::   |||   |||
Db       4080 TTCTCTAGACTTCAGCCAGACAAACCTCTCTCCAGAACTCAGTCAAACAAACCTTTCCCC 4139

Qy        438 MetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeu 457
                 :::||||||:::         |||   :::|||                  |||
Db       4140 AGCCCTCGGTCAGATGCCCCTTTCTCCAGACCCCAGCCA---------------TACAAC 4184

Qy        458 LysPheSerIleAlaArgProAlaIleAspLysGluLeu 470
                 ||||||      ||||||      ||||||   |||
Db       4185 CCTTTCTCT------AGACCTCAGCCAGACAAACCTCTC 4217

RESULT 11
US-08-454-353A-1
; Sequence 1, Application US/08454353A
; Patent No. 6518016
;  GENERAL INFORMATION:
;    APPLICANT:  Bertina, Rogier
;    APPLICANT:  Reitsma, Pieter
;    TITLE OF INVENTION:  A method for diagnosing an increased
```

```
;    TITLE OF INVENTION:  risk for thrombosis or a genetic defect causing
;    TITLE OF INVENTION:  thrombosis and kit for use with the same.
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Akzo No. 6518016el Patent Department
;      STREET:  1300 Piccard Drive, Suite 206
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/454,353A
;      FILING DATE:  06-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gormley, Mary E.
;      REGISTRATION NUMBER:  34,409
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 948-7400
;      TELEFAX:  (301) 948-9751
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6909 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-454-353A-1

Alignment Scores:
Pred. No.:              0.0681        Length:        6909
Score:                  110.50        Matches:       109
Percent Similarity:     37.12%        Conservative:  74
Best Local Similarity:  22.11%        Mismatches:    199
Query Match:            3.67%         Indels:        112
DB:                     4             Gaps:          28

US-09-913-878A-2_COPY_710_1282 (1-573) x US-08-454-353A-1 (1-6909)

Qy     28 GlyArgPhe------GlySerAlaLysGlyMetTrpValIle------AspValAsp--- 42
Db   2923 GGGAGATGGCATTTGGCTTCTGAGAAAGGTAGCTATGAAATAATCCAAGATACTGATGAA 2982

Qy     43 AspThrGlyAspGluAspTrp---IleGluThrTyrProSerGlnArgLysTrp------ 59
Db   2983 GACACAGCTGTTAACAATTGGCTGATCAGCCCCCAGAATGCCTCACGTGCTTGGGGAGAA 3042

Qy     60 GluCysAspPheValAspLys------------------HisGlnArgThrLeuGluVal 73
Db   3043 AGCACCCCTCTTGCCAACAAGCCTGGAAAGCAGAGTGGCCACCCAAAGTTTCCTAGAGTT 3102

Qy     74 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 93
Db   3103 AGACAT-------------------------------------AAATCTCTACAAGTAAGA 3126

Qy     94 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 113
Db   3127 CAGGATGGAGGAAAGAGTAGACTGAAGAAAAGCCAGTTT------------CTCATTAAG 3174

Qy    114 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu------ 131
Db   3175 ------ACACGAAAAAAGAAAAAAGAGAAGCACACACACCATGCTCCTTTATCTCCGAGG 3228

Qy    132 ------------PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgVal 147
Db   3229 ACCTTTCACCCTCTAAGAAGTGAAGCCTACAACACATTTTCAGAAAGA------AGACTT 3282

Qy    148 SerHisGlyArgValProPheLeuAla---------GlyLeuProAspSerGlnGluGlu 164
Db   3283 AAGCATTCGTTGGTGCTTCATAAATCCAATGAAACATCTCTTCCCACAGACCTCAATCAG 3342

Qy    165 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAsp 184
Db   3343 ACATTGCCC---TCTATGGATTTTGGCTGGATAGCCTCACTTCCTGACCATAATCAGAAT 3399

Qy    185 IleAlaTrpAspLeuGlnLysArgLysCys---------------AspThrLeuLysSer 199
Db   3400 TCCTCAAATGACACTGGTCAGGCAAGCTGTCCTCCAGGTCTTTATCAGACAGTGCCCCCA 3459

Qy    200 LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219
Db   3460 GAGGAACACTATCAAACA------------------------------------TTCCCC 3483

Qy    220 ValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGlu 239
Db   3484 ATTCAAGACCCTGATCAAATGCACTCTACTTCAGACCCCAGTCACAGATCCTCTTCTCCA 3543

Qy    240 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 259
Db   3544 GAGCTCAGTGAAATGCTTGAGTATGAC------------CGAAGTCACAAGTCCTTCCCC 3591

Qy    260 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 279
Db   3592 ACAGATATAAGTCAAATGTCCCCTTCCTCAGAA---------CATGAAGTCTGGCAGACA 3642

Qy    280 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyr 299
Db   3643 GTCATCTCTCCAGACCTCAGCCAGGTGACCCTCTCTCCAGAACTCAGCCAGACAAACCTC 3702

Qy    300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsn--- 318
Db   3703 TCTCCAGACCTCAGCCACACGACTCTCTCTCCAGAACTCATTCAGAGAAACCTTTCCCCA 3762

Qy    319 ------AlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThr 336
Db   3763 GCCCTCGGTCAGATGCCCATTTCTCCAGACCTCAGCCAT------------------ACA 3804

Qy    337 ThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThr 356
Db   3805 ACCCTTTCTCCAGACCTC---AGCCATACAACCCTTTCTTTAGACCTCAGCCAGACAAAC 3861

Qy    357 Tyr------AspMetIleGlnLysSerPhe------------------------------ 364
Db   3862 CTCTCTCCAGAACTCAGTCAGACAAACCTTTCCCCAGCCCTCGGTCAGATGCCCCTTTCT 3921

Qy    365 ------------HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
Db   3922 CCAGACCTCAGCCATACAACCCTTTCTCTAGACTTC------AGCCAGACAAACCTCTCT 3975

Qy    381 GluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIle--------- 397
Db   3976 CCAGAACTCAGCCATATGACTCTCTCT------------CCAGAACTCAGTCAGACAAAC 4023

Qy    398 LeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGlu 417
Db   4024 CTTTCCCCAGCCCTTGGTCAGATGCC-CATTTCTCCAGACCTCAG---CCATACAACCCT 4079

Qy    418 AlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspPro 437
Db   4080 TTCTCTAGACTTCAGCCAGACAAACCTCTCTCCAGAACTCAGTCAAACAAACCTTTCCCC 4139

Qy    438 MetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeu 457
Db   4140 AGCCCTCGGTCAGATGCCCCTTTCTCCAGACCCCAGCCA---------------TACAAC 4184

Qy    458 LysPheSerIleAlaArgProAlaIleAspLysGluLeu 470
Db   4185 CCTTTCTCT------AGACCTCAGCCAGACAAACCTCTC 4217

RESULT 12
US-09-165-019-1
; Sequence 1, Application US/09165019
```

```
; Patent No. 6558913
;  GENERAL INFORMATION:
;    APPLICANT:  Bertina, Rogier
;    APPLICANT:  Reitsma, Pieter
;    TITLE OF INVENTION:  A method for diagnosing an increased
;    TITLE OF INVENTION:  risk for thrombosis or a genetic defect causing
;    TITLE OF INVENTION:  thrombosis and kit for use with the same.
;    NUMBER OF SEQUENCES:  14
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/165,019
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/454,353
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6909 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-09-165-019-1

Alignment Scores:
Pred. No.:              0.0681         Length:         6909
Score:                  110.50         Matches:        109
Percent Similarity:     37.12%         Conservative:   74
Best Local Similarity:  22.11%         Mismatches:     199
Query Match:            3.67%          Indels:         112
DB:                     4              Gaps:           28

US-09-913-878A-2_COPY_710_1282 (1-573) x US-09-165-019-1 (1-6909)

Qy         28 GlyArgPhe------GlySerAlaLysGlyMetTrpValIle------AspValAsp--- 42
              |||||:::          |||  |||||   :::   |||        |||  |||
Db       2923 GGGAGATGGCATTTGGCTTCTGAGAAAGGTAGCTATGAAATAATCCAAGATACTGATGAA 2982

Qy         43 AspThrGlyAspGluAspTrp---IleGluThrTyrProSerGlnArgLysTrp------ 59
              ||||||          :::|||  |||                :::  |||  |||
Db       2983 GACACAGCTGTTAACAATTGGCTGATCAGCCCCCAGAATGCCTCACGTGCTTGGGGAGAA 3042

Qy         60 GluCysAspPheValAspLys-----------------HisGlnArgThrLeuGluVal 73
                          :::|||                   |||   :::         |||
Db       3043 AGCACCCCTCTTGCCAACAAGCCTGGAAAGCAGAGTGGCCACCCAAAGTTTCCTAGAGTT 3102

Qy         74 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 93
              |||                                      :::   |||   |||
Db       3103 AGACAT-------------------------------------AAATCTCTACAAGTAAGA 3126

Qy         94 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 113
              :::|||    :::   ::::::|||        |||             ||||||
Db       3127 CAGGATGGAGGAAAGAGTAGACTGAAGAAAAGCCAGTTT------------CTCATTAAG 3174

Qy        114 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu------ 131
                       |||:::      ::::::||||||      :::   |||:::
Db       3175 ------ACACGAAAAAAGAAAAAAGAGAAGCACACACACCATGCTCCTTTATCTCCGAGG 3228

Qy        132 ------------PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgVal 147
                             |||      |||   ::::::|||   |||      |||:::
Db       3229 ACCTTTCACCCTCTAAGAAGTGAAGCCTACAACACATTTTCAGAAAGA------AGACTT 3282

Qy        148 SerHisGlyArgValProPheLeuAla---------GlyLeuProAspSerGlnGluGlu 164
               |||     |||         :::           ||||||               :::
Db       3283 AAGCATTCGTTGGTGCTTCATAAATCCAATGAAACATCTCTTCCCACAGACCTCAATCAG 3342

Qy        165 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAsp 184
              ||||||         |||:::   |||:::                   :::   |||:::
Db       3343 ACATTGCCC---TCTATGGATTTTGGCTGGATAGCCTCACTTCCTGACCATAATCAGAAT 3399

Qy        185 IleAlaTrpAspLeuGlnLysArgLysCys---------------AspThrLeuLysSer 199
                 :::   |||      :::   |||                       |||:::
Db       3400 TCCTCAAATGACACTGGTCAGGCAAGCTGTCCTCCAGGTCTTTATCAGACAGTGCCCCCA 3459

Qy        200 LysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGly 219
              :::   :::   :::                                         :::
Db       3460 GAGGAACACTATCAAACA----------------------------------TTCCCC 3483

Qy        220 ValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGlu 239
              :::   :::   :::::::::|||                        |||
Db       3484 ATTCAAGACCCTGATCAAATGCACTCTACTTCAGACCCCAGTCACAGATCCTCTTCTCCA 3543

Qy        240 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 259
                 :::   :::   :::   |||                 ||||||         ||||||
Db       3544 GAGCTCAGTGAAATGCTTGAGTATGAC------------CGAAGTCACAAGTCCTTCCCC 3591

Qy        260 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 279
              :::||||||   ::::::                 :::         |||   :::
Db       3592 ACAGATATAAGTCAAATGTCCCCTTCCTCAGAA---------CATGAAGTCTGGCAGACA 3642

Qy        280 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyr 299
              :::|||                  |||   |||:::   :::||||||      :::
Db       3643 GTCATCTCTCCAGACCTCAGCCAGGTGACCCTCTCTCCAGAACTCAGCCAGACAAACCTC 3702

Qy        300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsn--- 318
                 |||::::::               ||||||::::::            ::::::
Db       3703 TCTCCAGACCTCAGCCACACGACTCTCTCTCCAGAACTCATTCAGAGAAACCTTTCCCCA 3762

Qy        319 ------AlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThr 336
                    :::||||||:::   |||||||||||||
Db       3763 GCCCTCGGTCAGATGCCCATTTCTCCAGACCTCAGCCAT-----------------ACA 3804

Qy        337 ThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThr 356
              |||         :::   ||||||   |||   |||            ||||||
Db       3805 ACCCTTTCTCCAGACCTC---AGCCATACAACCCTTTCTTTAGACCTCAGCCAGACAAAC 3861

Qy        357 Tyr------AspMetIleGlnLysSerPhe---------------------------- 364
                    ::::::   |||   :::
Db       3862 CTCTCTCCAGAACTCAGTCAGACAAACCTTTCCCCAGCCCTCGGTCAGATGCCCCTTTCT 3921

Qy        365 ------------HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLys 380
                          |||      |||         :::|||         ||||||
Db       3922 CCAGACCTCAGCCATACAACCCTTTCTCTAGACTTC------AGCCAGACAAACCTCTCT 3975

Qy        381 GluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIle--------- 397
                 |||      ::::::                  |||   :::
Db       3976 CCAGAACTCAGCCATATGACTCTCTCT------------CCAGAACTCAGTCAGACAAAC 4023

Qy        398 LeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGlu 417
              ||||||      :::|||   :::            |||:::            :::|||
Db       4024 CTTTCCCCAGCCCTTGGTCAGATGCC-CATTTCTCCAGACCTCAG---CCATACAACCCT 4079

Qy        418 AlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspPro 437
                 |||      |||      :::   |||               :::   |||   |||
Db       4080 TTCTCTAGACTTCAGCCAGACAAACCTCTCTCCAGAACTCAGTCAAACAAACCTTTCCCC 4139

Qy        438 MetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeu 457
                 :::||||||:::               |||   :::|||               |||
Db       4140 AGCCCTCGGTCAGATGCCCCTTTCTCCAGACCCCAGCCA---------------TACAAC 4184

Qy        458 LysPheSerIleAlaArgProAlaIleAspLysGluLeu 470
                 ||||||      ||||||   ||||||   |||
Db       4185 CCTTTCTCT------AGACCTCAGCCAGACAAACCTCTC 4217

RESULT 13
US-09-620-312D-985
; Sequence 985, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
;  APPLICANT:  Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
```

```
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  John Tillinghast
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. 6569662el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/09/620,312D
;  CURRENT FILING DATE:  2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1105
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 985
;   LENGTH: 2077
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (287)..(1912)
US-09-620-312D-985

Alignment Scores:
Pred. No.:               0.0218        Length:         2077
Score:                   107.50        Matches:        107
Percent Similarity:      31.75%        Conservative:   67
Best Local Similarity:   19.53%        Mismatches:     177
Query Match:             3.57%         Indels:         197
DB:                      4             Gaps:           26

US-09-913-878A-2_COPY_710_1282 (1-573) x US-09-620-312D-985 (1-2077)

Qy         19 LeuGlyAspValProSerAlaValGlnGly---------ArgPheGlySerAlaLysGly 35
              ||||||||              :::   ||        |||   ||||||
Db        572 TTGGGTGACCCAGAAGAACCGTTAACAGGATTTCCTGGAGAGGGGGATCT--------- 622

Qy         36 MetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyr------ 53
                         |||   :::   |||||   :::   |||   |||   :::
Db        623 -----------GATCCAGAAACCACTGGGATTCAAATCTGGAGTGAAGTTTTCACTGTG 670

Qy         54 ------ProSerGlnArgLysTrpGluCysAspPheValAspLysHis----------- 67
                    |||      :::|||         :::|||
Db        671 GAGAAGCCAGGTGGGAAGAAGGTTGCAGTTGTTCTGATGGATACCCAGGGGGCATTTGAC 730

Qy         68 ---GlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsn 86
                 |||   |||::::::      :::               |||   :::         :::
Db        731 AGCCAGTCAACTGTGAAAGACTGTGCTACCATCTTTGCTCTAAGCACTATGACTAGTTCT 790

Qy         87 LeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAla 106
              :::|||:::       :::   ::::::            |||   ::::::      |||
Db        791 GTTCAGATTTATAATTTATCTCAGAACATTCAAGAAGATGATCTTCAACAGCTGCAGCTC 850

Qy        107 IleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeu 126
                                                  |||:::|||
Db        851 ------------------------------------TTCACAGAA--------------- 859

Qy        127 AsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArg 146

Db        859 ------------------------------------------------------------ 859

Qy        147 ValSerHisGlyArgVal-----------------------ProPheLeuAlaGlyLeu 158
                    :::||||||:::                        ||||||
Db        860 ------TACGGTCGTCTGGCAATGGATGAAATTTTCCAAAAGCCTTTC----------- 901

Qy        159 ProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLys 178
                             :::||||||   ||||||:::               |||   :::
Db        902 --------------CAGACACTGATGTTTTTGGTTAGAGATTGGAGTTTCCCTTATGAA 946

Qy        179 GlnLysTyr---LeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeu 197
                    |||   ||||||               |||   ||||||
Db        947 TATAGCTATGGACTCCAAGGAGGAATGGCATTTTTGGATAAGCGT-------------- 991

Qy        198 LysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPhe 217
                       |||   ::::::      :::          |||         :::
Db        992 ---------TTACAGGTGAAGGAACATCAACATGAAGAAATTCAGAATGTT--------- 1033

Qy        218 TrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGlu 237
                                   |||   :::|||      ||||||
Db       1034 --------------CGAAATCACATTCACTCATGTTTCTCC------------------ 1060

Qy        238 GluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHis 257
              :::   :::         ||||||            :::   ||||||   ||||||
Db       1061 GATGTCACCTGCTTTCTCTTACCACATCCAGGACTCCAGGTGGCCACAAGCCCT---GAC 1117

Qy        258 PheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLys 277
              |||         ::::::   :::            ||||||   :::|||   :::|||
Db       1118 TTTGATGGGAAATTAAAAGATATTGCTGGTGAATTCAAAGAGCAGTTACAGGCACTGATA 1177

Qy        278 AspValIleIlePheSerThrLys----------------------------------- 285
                    ::::::      :::|||
Db       1178 CCGTATGTATTAAACCCATCTAAGTTAATGGAAAAGGAGATCAATGGCTCAAAGGTCACC 1237

Qy        286 -----------------------------------------------GlyAspVal 288
                                                              |||:::
Db       1238 TGTCGGGGACTACTGGAGTATTTTAAGGCATATATTAAAATTTATCAAGGAGAAGATCTG 1297

Qy        289 Pro----------------------------------------LeuAlaLysLysLeu 294
              |||                                         ||||||
Db       1298 CCTCACCCCAAGTCCATGCTTCAGGCCACTGCTGAAGCCTACAACTTAGCAGCTGCAGCC 1357

Qy        295 SerGlyGlyAsp---TyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleVal 313
              |||   |||   |||      :::|||
Db       1358 TCTGCCAAGGACATTTATTATAACAACATGGAAGAGGTTTGT----------------- 1399

Qy        314 AspGlyPheValAsnAlaGluMetPro---LeuGluProAspLeuSerArgTyrLeuLys 332
                          |||   |||   |||   ||||||:::         |||:::
Db       1400 -----------GGGGGAGAGAAACCTTATTTGTCTCCAGACATT---------CTAGAG 1438

Qy        333 LysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLys 352
              :::         ||||||||||||         |||
Db       1439 GAGAAGCACTGTGAATTCAAACAACTTGCTCTGGACCAT-------------------- 1477

Qy        353 GluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPhe 372
                 :::   |||         |||      |||   |||   |||         |||
Db       1478 TTTAAGAAGACCAAGAAGATGGGTGGGAAGGATTTCAGCTTTCGTTACCAGCAGGAGCTG 1537

Qy        373 LeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSer--- 391
                                |||   |||      :::      |||   |||   :::
Db       1538 GAGGAGGAAATCAAGGAATTATATGAGAACTTCTGCAAGCACAATGGTAGCAAGAACGTC 1597

Qy        392 -----------AsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAsp 407
                         ||||||::::::   :::   :::|||               :::
Db       1598 TTCAGCACCTTCCGAACCCCTGCAGTGCTGTTCACGGGCATTGTAGCTTTGTACATAGCC 1657

Qy        408 GlnSerLysGlnGlyIleValPheAsnGluAla------------------------- 418
                       |||   :::      |||
Db       1658 TCAGGCCTCACTGGCTTCATAGGTCTTGAGGTTGTAGCCCAGTTGTTCAACTGTATGGTT 1717

Qy        419 ----------------------SerTrpAlaGlnLeuArg---------------Arg 425
                                    :::|||   :::|||                |||
Db       1718 GGACTACTGTTAATAGCACTCCTCACCTGGGGCTACATCAGGTATTCTGGTCAATATCGT 1777

Qy        426 GluLeuLeuGlyGlyAlaLeuSerLeuProAspPro------MetTyrLysSerAspSer 443
              |||   ||||||||||||:::                     :::   ::::::   |||
```

```
Db       1778 GAG---CTGGGCGGAGCTATTGATTTGGTGCCGCATATGTGTTGGAGCAGGCTTCTTCT 1834

Qy        444 TrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArg 463
Db       1835 CATATCGGTAATTCCACTCAGGCCACTGTGAGGGATGCAGTT---------GTTGGAAGA 1885

Qy        464 ProAlaIleAspLysGluLeuGlu 471
Db       1886 CCATCCATGGATAAAAAAGCTCAA 1909

RESULT 14
US-09-489-039A-4533
; Sequence 4533, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 4533
;   LENGTH: 5307
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4533

Alignment Scores:•
Pred. No.:              0.101          Length:         5307
Score:                  107.50         Matches:        131
Percent Similarity:     33.48%         Conservative:   98
Best Local Similarity:  19.15%         Mismatches:     236
Query Match:            3.57%          Indels:         219
DB:                     4              Gaps:           33

US-09-913-878A-2_COPY_710_1282 (1-573) x US-09-489-039A-4533 (1-5307)

Qy          8 SerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal--------------- 22
Db       3043 ACAGTAGCTAAACAAGTCGATATAACCCTGAATGATGGCACAGTAGTTAAACAGTATGCA 3102

Qy         23 ---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspVal 41
Db       3103 TCTCCCCACCTCTGGAAGGGATATAGAGGACAGTCTACGTTGCCCTACTTATTACAGAGT 3162

Qy         42 AspAspThrGlyAspGluAspTrpIleGluThrTyr------------------ProSer 55
Db       3163 GCTCTTATGGCATTGGAGAACTGGTTAGTTGAGTACGTTGAAAATTGTGGTAAGAATAAT 3222

Qy         56 GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSer 75
Db       3223 GAAATAGATTGGATTTTGATTATGTTCTTCGCTCTAGTAATTCAGTAATGCCTACATCT 3282

Qy         76 ValAlaSerGluLeuLysSerAlaGlyLeuAsn------------------LeuGlnLeu 89
Db       3283 GTTTTATCTTCTGTGGCCACAGGCTTCCCTAATAAAGTAGGTAAAGCAGCCTTTCCTTTG 3342

Qy         90 LeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAsp 109
Db       3343 CTTAAAACTGCTGATTTATACCATTTAGATTTGATTCGAATGACACAAGAAATGGGTGGA 3402

Qy        110 ArgLeuIle---------------------AsnAspLeuGlnArgGlnPheSerGluGln 122
Db       3403 AATGAGATGCATTTCTTTGGTCTTAATAGGGATGCATTGTCGAAAATATACCTAGAAGAG 3462

Qy        123 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 142
Db       3463 AGA------------AGAGAAGCAGCACTTCGACCATGGAGAAAAGAGTCATTAGAAACC 3510

Qy        143 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp------ 160
Db       3511 TTATTAACAAGGCTTCAG---------------TTCGTAAATGAGTTAAGGGATGATATT 3555

Qy        161 ---------------------------------------SerGlnGluGluThrLeuAsn 167
Db       3556 TTAAAAATTGTTGATGAACTGAAAAATGAAGCTACTGCAAGTAATGAAAAAAGCCTTCGG 3615

Qy        168 PheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAsp--------- 184
Db       3616 TATATGGTGCACAGA---GTCGATACAAGGACTTGGAAAGTAGTTGAAGACAAAGAAAAT 3672

Qy        185 -----------------------------IleAlaTrpAspLeu---------------- 189
Db       3673 GACAGAGTATTGTTTCAAAGCTCATCAGAGCTTCCGGAAGATTTAAAGCTGGATCAACAA 3732

Qy        190 ---------GlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGly 206
Db       3733 GAATTTAATGAAAAGCACGCAATGGATAACACCGTAACTAGTTTAAACTTA--------- 3783

Qy        207 ArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluVal 226
Db       3784 ---------------------------------TGGGGT--------------------- 3789

Qy        227 HisValGlyPheSerSerLysPheArgAspGluGlu------GluSerPheThrLeuLeu 244
Db       3790 ---------------AAAAAACTATTCGATGAGCAATTACTAGAAGAAAAATACTTCTCT 3834

Qy        245 SerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArg 264
Db       3835 TCATATCAGGATGCATTAATAGCTACTAAGGGATTG------------------------ 3870

Qy        265 ValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIle------Phe 282
Db       3871 CTCAGCGCCCTCCAAAAAGGTGAAGTACACAACTTTGCTGATATGGCGGTTGGCACAATC 3930

Qy        283 SerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAsp 302
Db       3931 ACAACTGTTGCAGCGGTTTGTGTTCGTGATGTTCTTTTAGAATTGAGCTATGAA---GAT 3987

Qy        303 MetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetPro 322
Db       3988 AAAAAGTGG---TGCTTAGAAATCATCCTAGAATCAATATTCATGCACGCAGAC------ 4038

Qy        323 LeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMet 342
Db       4039 ---------AATATGAATGGGACAGCTGCACATGATAAGACTGATTAT------------ 4077

Qy        343 AlaSerHisGlyThrGlySerAlaAla--------------------------------- 351
Db       4078 ------TATGGCTCTGGAGCATGCGCATTTGTCTTACCTAAGTTATTTGATTTAGATCTA 4131

Qy        352 ---LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 370
Db       4132 GATTCAGAACAAATAGAATAT------------------CTGAAGTTTGCGTTA------ 4167

Qy        371 AsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerVal 390
Db       4168 ---------------------------GCTACAGCATTAACGCATGAGAACCTGAACGTT 4200

Qy        391 Ser-------------------------------------AsnLysProAlaIleIleLeu 398
Db       4201 AGTGCTTGTGCTGCTAAGGGAGTAAGAGAGTTTCTTTGGTCTAGAGACGCTGAATTAGCA 4260

Qy        399 SerSerLeuValGlyAsnLeuValAspGlnSerLys------------------------ 410
Db       4261 TCCCGTTGTATAGGTGGAATAGTTGAGTATGCCAGATTACGAAGGGAAGATAGTGAGGTA 4320

Qy        411 ------------------GlnGlyIleValPheAsnGluAla------SerTrpAlaGln 422
Db       4321 CGTAGGTTTTATCATTTACAAGGTGCTGAATTGCAAGCCGCACTCGAAAAATGGAATGAT 4380

Qy        423 Leu------------ArgArgGluLeuLeuGlyGly------AlaLeuSerLeuProAsp 436
```

```
Db       4381 CTTTTAACAAGTTTTCGTAACGACTTGATTGAGGGAAAGTTTAAGCTTTCAGTTAACGAT 4440

Qy        437 ProMetTyrLysSerAsp---SerTrpLeuGlyArgGlyGluProThrHisIleIleAsp 455
                         :::|||    |||||           |||      ::::::
Db       4441 ATATCTTTGGAAAGCCACTGTTCGTGGTTTATTCAT---TTACCAATGCTAATGGTTCCT 4497

Qy        456 TyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsn 475
              |||                            :::  :::::::::    :::
Db       4498 TACTGTACTAAGGAT------------------GAGATACAAATTCAGTTGGTCAATAGG 4539

Qy        476 AlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAla 495
                 |||          ||||||||        |||          |||
Db       4540 TTAATGAGTTTTGTATTTGATACGGAGTATCAAAACCATCGTTCGAGTGATGATGAAAAA 4599

Qy        496 SerTyrTyrThrPhePheLysGluIleSerAsp--------------------LysSer 508
                    :::          |||:::|||  |||                    |||
Db       4600 ATAAATCATGAAATTAAAAAGCAGATTCAAGACTGTCTCTCTGAACATGTTGTTTCTTCA 4659

Qy        509 ArgSerSer-------------------------AlaLeuLeuPheThrThrLeuLys 519
              |||   :::                                |||      |||
Db       4660 AGAAAAAATAACTTCATGCCCTTTAAAGATCTATTGGTGTTAGGGTGTACAAAAGCTCCC 4719

Qy        520 AsnArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGluMetArg 539
                    |||        :::         |||         |||:::   :::
Db       4720 GGCTTCATTTATTCTTTAACACTCTCCTATCATGTCGCTGTAGAAAAAGAAGGTGATTAT 4779

Qy        540 AspSerLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIle 559
              |||:::                                ::::::     |||||    :::
Db       4780 GATGCT----------------------------ATATGGGCACTTTGGTGCCTGTTA 4809

Qy        560 ThrProGluAla 563
                 |||||||||
Db       4810 GCACCAGAAGCA 4821
```

RESULT 15
US-09-073-492-1
; Sequence 1, Application US/09073492
; Patent No. 6248578
; GENERAL INFORMATION:
;   APPLICANT: Banerjee, Amiya K.
;   APPLICANT: Hoffman, Michael A.
;   TITLE OF INVENTION: Infectious Clone for Human Parainfluenza
;   TITLE OF INVENTION: Virus Type 3
;   NUMBER OF SEQUENCES: 5
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Calfee, Halter & Griswold LLP
;   STREET: 800 Superior Avenue
;   CITY: Cleveland
;   STATE: Ohio
;   COUNTRY: USA
;   ZIP: 44114-2688
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/073,492
;   FILING DATE:
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Docherty, Pamela A.
;   REGISTRATION NUMBER: 40,591
;   REFERENCE/DOCKET NUMBER: 23114/04000
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (216)622-8200
;   TELEFAX: (216)241-0816
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 15462 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
US-09-073-492-1

```
Alignment Scores:
Pred. No.:                 1               Length:         15462
Score:                     105.50          Matches:        126
Percent Similarity:        35.09%          Conservative:   100
Best Local Similarity:     19.57%          Mismatches:     213
Query Match:               3.50%           Indels:         205
DB:                        3               Gaps:           34
```

US-09-913-878A-2_COPY_710_1282 (1-573) x US-09-073-492-1 (1-15462)

```
Qy         40 AspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrp 59
              |||:::||||||      |||          :::   |||         ::::::|||
Db       8772 GATATGGATGAC------GACTCAATACTAGTTATCACT---------AGACAGAAAATA 8816

Qy         60 GluCysAspPheValAspLysHisGlnArgThrLeu----------------------- 71
              :::   :::   :::||||||   ||||||::::::
Db       8817 AAACTCAATAAATTAGATAAAAGACAACGATCTATTAGAAGATTAAAATTAATATTAACT 8876

Qy         72 --------------------------------------------GluValArgSerVal 76
                                                          |||:::
Db       8877 GAGAAAGTGAATGACTTAGGAAAATACACATTTATTAGATATCCAGAAATGTCAAAAGAA 8936

Qy         77 AlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArg 96
                    :::|||          |||:::|||   ::::::         :::|||
Db       8937 ATGTTCAAATTACATATACCTGGTATTAACAGTAAAGTGACTGAATTATTACTTAAAGCA 8996

Qy         97 AlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGln 116
                 |||          :::|||         :::   |||      ||||||   |||
Db       8997 GATAGAACATATAGTCAAATGACTGATGGATTAAGAGATCTATGGATTAATGTGCTATCG 9056

Qy        117 ArgGlnPheSerGlu---------GlnLysHisAlaLeuAsnArgProValGlu------ 131
              :::      |||:::               :::   ||||||         :::
Db       9057 AAATTAGCCTCAAAAAATGATGGAAGCAATTATGATCTTAATGAAGAAATTAATAATATA 9116

Qy        132 --------------------------------------------PheArgGlnTrpVal 136
                                                          |||:::   |||
Db       9117 TCAAAAGTTCACACAACCTATAAATCAGATAAATGGTATAATCCATTCAAAACATGGTTC 9176

Qy        137 TyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAla 156
                       |||      |||      :::   ::::::         |||   |||
Db       9177 ACTATCAAGTATGATATGAGAAGATTGCAAAAAGCTCGAAATGAGGTCACTTTTAATATG 9236

Qy        157 GlyLeuPro---AspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAsp 175
              |||         :::      |||:::      ||||||||:::
Db       9237 GGGAAAGATTATAACTTGTTAGAAGACCAGAAGAATTTCTTATTG---------ATACAT 9287

Qy        176 ProLys------------------LysGlnLysTyrLeu---GlnAspIleAlaTrpAsp 188
              |||:::                  ||||||   |||         |||         :::
Db       9288 CCAGAATTGGTTTTAATATTAGATAAACAAAACTATAATGGTTATCTAATTACTCCTGAA 9347

Qy        189 LeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgVal--------- 205
              |||               ||||||   ::::::   :::   ||||||
Db       9348 TTAGTATTGCCGTATTGTGACGTAGTTGAAGGCCGATGGAATATAAGTGCATGTGCTAAG 9407

Qy        206 ------GlyArgSerAlaTyrIleTyrMetIleAla---AspPheTrpGlyValLeuGlu 222
                          :::         :::|||            :::   |||   |||::::::
Db       9408 TTAGATCCAAAATTACAATCTATGTATCAGAAAGGCAATAATCTGTGGGAAGTGATAGAT 9467

Qy        223 GluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheThr 242
              :::            :::|||                              |||::::::|||
Db       9468 AAATTGTTTCCAATTATGGGA------------------------GAAAAGACATTTGAT 9503

Qy        243 LeuLeuSerAspCysAspValLeu---ValAlaArgSerProAlaHisPheProSerAsp 261
              ::::::|||      :::   |||      ::::::            |||   |||
Db       9504 GTGATATCATTATTAGAACCACTTGCATTATCTCTAATTCAAACTCATGATCCT------ 9557

Qy        262 IleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal---Ile 280
              ::::::::::::|||         |||                  |||   |||   ::::::   :::
```

```
Db       9558 GTTAAACAATTAAGGGGAGCTTTTTAAAT------CATGTGTTATCCGAGATGGAATTG 9611

Qy        281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 300
              ||||||   ::::::     :::          |||:::          ||||||
Db       9612 ATATTTGAATCTAGAGAATCGATT---------AAAGAATTTCTGAGTGTAGATTAC--- 9659

Qy        301 GlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGlu 320
                                                :::|||:::|||   |||      :::
Db       9660 -----------------------ATTGATAAAATCTTAGATATATTTAATAAATCTACA 9695

Qy        321 MetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPhe------ 338
              :::       ::::::    :::   :::                ||||||
Db       9696 ATAGATGAAATAGCAGAGATTTTCTCTTTTTTAGA-----------ACATTTGGGCAT 9743

Qy        339 LysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLys--------------Glu 353
                    |||  ||||||                      |||                :::
Db       9744 CCTCCATTAGAGGCTAGTATTGCAGCAGAAAAAGTTAGAAAATATATGTATATTGGGAAA 9803

Qy        354 GlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeu 373
              |||      :::|||   |||   |||
Db       9804 CAATTAAAATTTGACACTATTAATAAATGTCATGCTATC--------------------- 9842

Qy        374 GlyMetCysThr------------AsnTyrLysGluArgLeuCys-------------- 384
                    ||||||                  |||:::||||||
Db       9843 ---TTCTGTACAATAATAATTAACGGATATAGAGAAAGGCATGGTGGACAGTGGCCTCCT 9899

Qy        385 --------------------------TyrIleAsnAsnSerValSerAsnLysProAla 395
                                        :::|||   |||:::      :::      |||
Db       9900 GTGACATTACCTGATCATGCACACGAATTCATCATAAATGCTTACGGTTCAAATTCTGCG 9959

Qy        396 IleIleLeuSerSerLeuValGlyAsnLeuValAsp------GlnSerLysGlnGlyIle 413
              |||                     |||   ||||||      ||||||     ||||||
Db       9960 ATATCATATGAA------------AACGCTGTTGATTATTACCAGAGCTTTATAGGAATA 10007

Qy        414 ValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSer 433
                 ||||||:::           ||||||       :::|||             :::
Db      10008 AAATTTAATAAATTCATAGAACCTCAGTTAGATGAAGATTTG---------ACAATTTAT 10058

Qy        434 LeuProAspProMetTyr------LysSerAspSerTrp-------------------- 444
              :::   |||                |||      :::|||
Db      10059 ATGAAAGATAAAGCATTGTCTCCAAAAAAAATCAAACTGGGACACAGTTTCTCCTGCATCT 10118

Qy        444 ------------------------------------------------------------ 444

Db      10119 AATTTACTGTACCGTACTAACGCATCCAACGAATCACGAAGATTAGTTGAAAAATTTATA 10178

Qy        445 ---LeuGlyArgGlyGluProThrHisIleIleAspTyrLeu----------------- 457
                       :::   :::|||         |||:::||||||:::
Db      10179 GCAGATAGTAAATTTGATCCTAATCAGATATTAGATTATGTAGAATCTGGGGACTGGTTA 10238

Qy        458 ---------LysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHis 474
                       :::|||:::|||:::          :::||||||::::::
Db      10239 GATGATCCAGAATTTAATATTTCTTATAGTCTTAAAGAAAAAGAGATCAAA--------- 10289

Qy        475 AsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeu 494
                                        ::::::|||
Db      10290 --------------------------CAAGAAGGT------------------------ 10298

Qy        495 AlaSerTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeu 514
                          |||   :::::::::   |||   |||::::::   :::|||
Db      10299 -----------AGACTCTTTGCAAAAATGACATATAAAATGAGAGCTACACAAGTTTTA 10346

Qy        515 Phe---ThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuVal 533
                    |||   |||   |||   ||||||:::      :::|||   |||   :::
Db      10347 TCAGAGACACTACTTGCAAATAATATAGGGAAATTCTTTCAAGAAAATGGGATGGTAAAA 10406

Qy        534 LysAsnLysGluMetArgAspSerLysAspProTyrProVal-----------ArgVal 549
                       |||:::                      :::               |||
Db      10407 GGAGAGATTGAATTACTTAAGAGATTAACAACCATATCAATATCAGGAGTTCCACGGTAT 10466

Qy        550 AsnGlnValTyr 553
              |||:::||||||
Db      10467 AATGAAGTATAC 10478
```

Search completed: March 31, 2004, 13:13:05
Job time : 138.859 secs

GenCore version 5.1.6


```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         March 31, 2004, 06:26:16 ; Search time 2230.2 Seconds
                                           (without alignments)
                                           7672.409 Million cell updates/sec

Title:          US-09-913-878A-2_COPY_710_1282
Perfect score:  3010
Sequence:       1 GVGRMSRSVAKRIRDVLGLG..........EKWCAITPEAMDKSGANYDS 573

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters:      55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09913878/runat_29032004_082008_8379/app_query.fasta_1.2254
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878_@CGN_1_1_4565_@runat_29032004_082008_8379 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      EST:*
                1:  em_estba:*
                2:  em_esthum:*
                3:  em_estin:*
                4:  em_estmu:*
                5:  em_estov:*
                6:  em_estpl:*
                7:  em_estro:*
                8:  em_htc:*
                9:  gb_est1:*
                10:  gb_est2:*
                11:  gb_htc:*
                12:  gb_est3:*
                13:  gb_est4:*
                14:  gb_est5:*
                15:  em_estfun:*
                16:  em_estom:*
                17:  em_gss_hum:*
                18:  em_gss_inv:*
                19:  em_gss_pln:*
                20:  em_gss_vrt:*
                21:  em_gss_fun:*
                22:  em_gss_mam:*
                23:  em_gss_mus:*
                24:  em_gss_pro:*
                25:  em_gss_rod:*
                26:  em_gss_phg:*
                27:  em_gss_vrl:*
                28:  gb_gss1:*
                29:  gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	240.5	8.0	2199	11	AY103827	AY103827 Zea mays
	2	204	6.8	883	14	CA765940	CA765940 AF53-Rpf_
	3	191.5	6.4	577	10	BF096404	BF096404 EST360431
c	4	184.5	6.1	954	29	CG445102	CG445102 OG8AH57TV
	5	183.5	6.1	641	14	CA257764	CA257764 SCCCFL110
	6	175.5	5.8	582	12	BI246074	BI246074 IP1_68_D1
	7	173	5.7	622	13	BU965311	BU965311 sat08g07.
	8	172	5.7	761	10	BF275078	BF275078 GA__Eb002
	9	171	5.7	934	29	CG212111	CG212111 OGOFJ51TV
	10	169.5	5.6	648	14	CA223511	CA223511 SCJFFL1C0
	11	169.5	5.6	741	14	CF441078	CF441078 EST677423
	12	168	5.6	541	13	BQ827639	BQ827639 gd58f03.y
	13	168	5.6	677	14	CD928156	CD928156 GR45.104C
	14	167.5	5.6	916	29	CG223171	CG223171 OGWAL14TV
	15	166	5.5	722	14	CF134248	CF134248 WHE4369_F
	16	166	5.5	812	29	CG140649	CG140649 PUIHY65TB
	17	164	5.4	570	10	BE433646	BE433646 EST400175
	18	163.5	5.4	585	14	CF244986	CF244986 3530_1_8_
	19	163	5.4	640	10	BF460205	BF460205 069G02 Ma
c	20	159.5	5.3	741	13	BQ997497	BQ997497 QGG16J23.
	21	157.5	5.2	621	13	BU997627	BU997627 HI08I16r
	22	154.5	5.1	629	12	BJ268864	BJ268864 BJ268864
c	23	154	5.1	781	29	CG938103	CG938103 MBENF75TR
	24	153.5	5.1	885	28	BH163571	BH163571 ENTSB82TF
c	25	151	5.0	658	28	BZ336439	BZ336439 hz34c12.b
c	26	150	5.0	352	14	CF053563	CF053563 QCN18d04.
c	27	145	4.8	351	14	CF053707	CF053707 QCN1e10.y
	28	144.5	4.8	967	14	CA272420	CA272420 SCVPLB209
	29	143.5	4.8	411	10	BF252757	BF252757 EST420020
c	30	143	4.8	342	14	CF055747	CF055747 QCN9g09.y
	31	143	4.8	683	14	CA199990	CA199990 SCRFFL103
	32	140.5	4.7	1193	14	CK210907	CK210907 FGAS02273
	33	140	4.7	442	28	BH739133	BH739133 BOHWT59TF
	34	140	4.7	725	12	BG594667	BG594667 EST493345
	35	139	4.6	295	28	BZ992509	BZ992509 PUGGQ01TD
	36	139	4.6	613	14	CA220413	CA220413 SCSBFL401
	37	138.5	4.6	597	10	AW649811	AW649811 EST328265
	38	138.5	4.6	895	12	BG344709	BG344709 HVSMEg001
	39	136.5	4.5	525	10	AW933692	AW933692 EST359535
	40	136	4.5	405	10	BF459767	BF459767 064G04 Ma
	41	134.5	4.5	517	12	BM368419	BM368419 EBem08_SQ
	42	131	4.4	554	14	CD896715	CD896715 G174.103L
	43	130.5	4.3	734	14	CF449332	CF449332 EST685677
	44	128.5	4.3	610	10	BE920043	BE920043 EST423812
	45	128	4.3	222	14	CA902339	CA902339 PCSC13733

ALIGNMENTS

```
RESULT 1
AY103827
LOCUS       AY103827                2199 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO135426 mRNA sequence.
ACCESSION   AY103827
VERSION     AY103827.1  GI:21206905
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2199)
```

```
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 2199)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..2199
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:637927"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN

Alignment Scores:
Pred. No.:              3.14e-17        Length:         2199
Score:                  240.50          Matches:        137
Percent Similarity:     36.87%          Conservative:   96
Best Local Similarity:  21.68%          Mismatches:     258
Query Match:            7.99%           Indels:         141
DB:                     11              Gaps:           23

US-09-913-878A-2_COPY_710_1282 (1-573) x AY103827 (1-2199)

Qy          1 GlyValGlyArgMetSerArgSerValAla--------------------------Lys 11
Db        171 GGAGTTGGAAAGATCTCAGCTAATTTGCAGTGGAGGTGGCTATGAAGTGCAAATTGAAA 230

Qy         12 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGly 31
Db        231 CGCTTT-----------------------GCTCCTTCTGTTTTCAGATAAGGTATGGC 266

Qy         32 SerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGlu 51
Db        267 GGTTACAAAGGTGTTGTCGCTGTA---------GATACAAGATCAAATCATAAGCTTTCT 317

Qy         52 ThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 71
Db        318 TTGAGAAAAAGCATGTCAAAGTTCCAGTCAGAAAATATC---------------ACTCTT 362

Qy         72 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuPro 91
Db        363 GATGTCCTTGCATAC---AGCAAGTACCAACCATGCTTCCTGAATCGGCAGTTGATTACT 419

Qy         92 ValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeu 111
Db        420 CTTCTCTCA--------------------------ACACTTGGGGTTAGCGATAATGTC 452

Qy        112 IleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu 131
Db        453 TTTGAGCTAAAGCAGAAGGAAGCCTTAAGGCAG---------TTGAACAGAATGGTAACT 503

Qy        132 PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArg 151
Db        504 ---------------------------GAACCACAGGCTGCTCGTGAAGCAGTTGAACTT 536

Qy        152 ValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsn 171
Db        537 ATGCCCATG---------------GGAGAAGTAACCAATGTAGTTAAAGAATTGTTGTCA 581

Qy        172 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 191
Db        582 TGTGGCTACCAGCCTGATCATGAGCCATATCTTTCCATGCTGCTACAAACTTTTAGAGCA 641

Qy        192 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 211
Db        642 TCCAAGCTTCTAGAGTTGAAAACAAAGTCAAGGATATTCATCACACAGGGGCGAGCAATG 701

Qy        212 TyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer 231
Db        702 ATGGGTTGCCTGGATGAAACCTGCACACTTAAGTACGGCCAGGTATTCGTCCAAGCTTCT 761

Qy        232 SerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 251
Db        762 TACAGTGCAGATGACCATCGCAAG---------GTCGTTGTAACTGGAAAAGTAGTTGTC 812

Qy        252 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 271
Db        813 GCCAAAAATCCTTGTCTCCACCCTGGTGACATACGGGTTCTCCAGGCTGTTGATGTTCCT 872

Qy        272 GluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAla 291
Db        873 GCTCTGCACCACTTGTTTGACTGTGTTGTCTTTCCACAGCAGGGACCAAGGCCGCACCCT 932

Qy        292 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 311
Db        933 AATGAGTGTTCAGGGAGTGATCTTGATGGGGACATATATTTGTTTCTTGGGATCCACAT 992

Qy        312 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu 331
Db        993 CTTATTCCAAGTCGTTTGGTGGATCCTATGGACTATACTCCAGCTTCAGCAGAAACATTA 1052

Qy        332 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 351
Db       1053 GACCATGAT--------------------------------------------------- 1061

Qy        352 LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 371
Db       1062 ---------GTCACTATTGAGGAGATACAGGAGTACTTCACAAACTACATAGTTAATGAG 1112

Qy        372 PheLeuGlyMetCysThrAsnTyr-----------------LysGluArgLeuCysTyr 385
Db       1113 AGTCTTGGGATTATCGCCAATGCGCATGTGGTCTTTACAGATCAGGAACGTATG------ 1166

Qy        386 IleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeu 405
Db       1167 ---------AAAGCTGAGAGTCCACCGTGCGTTCAACTGGCCAAGCTCTTCTCTATAGCT 1217

Qy        406 ValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArg 425
Db       1218 GTCGATTTCCCAAAGACTGGAGTGCCG-----------------GCTCTGATTCCACAT 1259

Qy        426 GluLeuLeuGlyGlyAlaLeuSerLeuProAspProMet-------------------- 438
Db       1260 GAGCTA------CATGTCAAGGAGTATCCTGACTTCATGGAGAAACTCGACAAAGTCACC 1313

Qy        439 TyrLysSerAspSerTrpLeuGlyArg---------------GlyGluProThrHisIle 453
Db       1314 TATGAATCAAAGGGTGTGATCGGGAAGCTCTATAGGGAAATAAAGAAGCACACACCACAC 1373

Qy        454 IleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeu------Glu 471
Db       1374 ATAAAGCACTTCACGAGGGAAGTGGCAAGGCGGTCTTATGACACCGATTTGATTGTTGAT 1433

Qy        472 AlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAsp 491
Db       1434 GGCTATGAAGATTACATTACTGAGGCTATAGAGTTCAAGGAAGAGTACGATTTCAGGCTG 1493
```

```
Qy        492 ProAspLeuAlaSerTyrTyrThrPhePheLysGlu---------IleSerAspLysSer 508
                 :::|||    :::|||                 |||      ||||||
Db       1494 GGTAATCTTATGGACCACTATGGCATAAAAGTGAAGCTGAGATAATAAGTGGATGTATT 1553

Qy        509 ArgSerSerAlaLeuLeuPheThr------------------ThrLeuLysAsnArgIle 522
                       |||     ||||||                     ::::::    :::
Db       1554 CTAAAGATGGCAAAGAATTTCACCAAGAGTAGTGATGCTGATGCAATTAGAATGGCGGTG 1613

Qy        523 GlyGluValGluLysGlu------------TyrGlyArgLeuValLysAsnLysGluMet 538
                    :::   ||||||              :::     :::      ::::::
Db       1614 AGATCTTTGAGGAAAGAAGCTAGGTCGTGGTTCAATGAGATGAGCACAGGAGAGGATGGC 1673

Qy        539 ArgAspSerLysAspPro---------------TyrProValArgValAsnGlnValTyr 553
               :::|||:::   :::                 |||   |||      :::|||   |||
Db       1674 CAAGATGCCATGGAGGCCAAGGCCTCTGCTTGGTACCATGTTACTTATCATCAGCAGTAC 1733

Qy        554 ---------------------------------------------GluLysTrpCysAla 558
                                                                 ||||||
Db       1734 TGGGGCAGCTACAATGAAGGGTATGATCGGCCGCATCTTATTAGCTTCCCATGGTGCGTA 1793

Qy        559 IleThrProGluAlaMetAspLysSerGlyAlaAsn 570
                                |||   |||   |||
Db       1794 TATGACAAGCTTGTGGCCATCAAGCAGGGGAGGAAT 1829
```

```
RESULT 2
CA765940
LOCUS       CA765940                 883 bp    mRNA    linear   EST 08-JAN-2003
DEFINITION  AF53-Rpf_09_H23_T7_092.ab1 IRRI Drought Stress Panicle Library
            Oryza sativa (indica cultivar-group) cDNA clone C0003263 5' similar
            to unknown, mRNA sequence.
ACCESSION   CA765940
VERSION     CA765940.2  GI:27547887
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 883)
  AUTHORS   Bennett,J., Arumugam,K., Lafitte,R., Wen,J., Rudd,S. and
            Bruskiewich,R.M.
  TITLE     IRRI Drought Stress Panicle cDNA Library
  JOURNAL   Unpublished (2002)
COMMENT     On Dec 2, 2002 this sequence version replaced gi:25995195.
            Contact: Richard Bruskiewich
            Biometrics and Bioinformatics Unit
            International Rice Research Institute
            DAPO 7777, Metro Manila, Philippines
            Tel: +63-2-845-0563
            Fax: +63-2-845-0606
            Email: r.bruskiewich@cgiar.org
            International Rice Information System (IRIS;
            http://www.iris.irri.org): D0203262
            Assignment of putative function to the sequence by S. Rudd of the
            Munich Information Center for Protein Sequences
            (http://mips.gsf.de)
            Plate: 09  row: H  column: 23.
FEATURES             Location/Qualifiers
     source          1..883
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="IR64"
                     /db_xref="taxon:39946"
                     /clone="C0003263"
                     /tissue_type="Panicles"
                     /dev_stage="Flowering"
                     /clone_lib="IRRI Drought Stress Panicle Library"
                     /note="Vector: pBluescript II SK+; Water stress was
                     applied by not watering for 4 consecutive days. Panicles
                     were collected from control (well watered) and stressed
                     plants at 2 days before heading, at heading, 50% flowering
                     and 4 days after 50% flowering."
ORIGIN

Alignment Scores:
Pred. No.:              1.96e-13        Length:          883
Score:                  204.00          Matches:         56
Percent Similarity:     46.26%          Conservative:    43
Best Local Similarity:  26.17%          Mismatches:      99
Query Match:            6.78%           Indels:          16
DB:                     14              Gaps:            3

US-09-913-878A-2_COPY_710_1282 (1-573) x CA765940 (1-883)

Qy        115 LeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGln 134
              |||   ||||||                       :::||||||   |||
Db         45 TTAAATCGACAGATTATAATATTGCTATCCTCACTGAATGTCCCAGATTCTATCTTTTGG 104

Qy        135 TrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPhe 154
                 :::   |||:::      |||            :::      |||
Db        105 CAAATGCAAGAGACCATGCTTTCTAACCTCAACAATATTCTATCAGACAGAGATGTTGCT 164

Qy        155 LeuAlaGlyLeuProAspSer---------GlnGluGluThrLeuAsnPheLeuMetAsn 171
                       |||      |||               |||        :::::::::
Db        165 TTTGAGGTTTTAACAACTTCATGTGCTGATGATGGAAACACTGCAGCATTGATGCTCAGT 224

Qy        172 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 191
              :::||||||:::|||:::   :::   :::|||:::         ::::::
Db        225 GCTGGCTTTGAACCTAGAACTGAACCACACTTGAAAGCAATGCTCTTGGCTATAAGGTCT 284

Qy        192 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 211
                 :::      |||   |||   |||   |||   :::      ::::::
Db        285 GCACAATTGCAGGATCTTTTGGAAAAAGCAAGGATATTTGTGCCAAAGGGAAGGTGGTTG 344

Qy        212 TyrMetIleAlaAspPheTrpGlyValLeuGluGlu------------------------ 223
                       |||      |||||||||||:::
Db        345 ATGGGCTGTCTTGATGAGCTTGGGGTTCTTGAGCAAGGGCAGTGCTTTATTCGGGCAACA 404

Qy        224 ------------AsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGlu 239
                          |||      |||      |||:::|||         ::::::
Db        405 GTTCCATCATTGAATAGTTATTTTGTTAAGCATGGGTCAAGATTTTCATCAACAGATAAA 464

Qy        240 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 259
              :::      ::::::            |||::::::|||::::::|||         |||
Db        465 AACACAGAGGTCATTTTGGGA---ACTGTGGTAATAGCAAAGAATCCCTGTCTTCATCCA 521

Qy        260 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 279
                 |||::::::   :::   ||||||      ||||||||||||   |||   |||
Db        522 GGGGATGTCCGCATCCTTGAAGCAGTTGATGTGCCCGAACTGCATCATCTGGTTGATTGT 581

Qy        280 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyr 299
              ::::::|||      ||||||:::   |||   |||      :::   ||||||   |||
Db        582 TTGGTGTTCCCCCAGAAAGGTGAGAGGCCACATGCTAACGAGGCATCTGGGAGCGATCTT 641

Qy        300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleVal 313
              |||||||||:::   :::|||   ||||||   :::::::::
Db        642 GATGGGGATCTCTACTTTGTGACATGGGATGAGAAACTTATA 683

RESULT 3
BF096404
LOCUS       BF096404                 577 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST360431 tomato nutrient deficient roots Lycopersicon esculentum
            cDNA clone cLEW11O6 5' sequence similar to RNA-directed RNA
            polymerase {Arabidopsis thaliana}, mRNA sequence.
ACCESSION   BF096404
VERSION     BF096404.1  GI:10902114
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (tomato)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   van der Hoeven,R.S., Garvin,D.F., Matern,A.L., Holt,I.E., Liang,F.,
            Upton,J., Hansen,T.S., Ronning,C.M., Craven,M.B., Bowman,C.L.,
```

```
            Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J.
            and Tanksley,S.D.
  TITLE     Generation of ESTs from tomato nutrient-deficient roots
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1..577
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA492"
                     /db_xref="taxon:4081"
                     /clone="cLEW11O6"
                     /tissue_type="roots"
                     /dev_stage="5-6 weeks old"
                     /lab_host="SOLR"
                     /clone_lib="tomato nutrient deficient roots"
                     /note="Vector: pBluescriptSKmCUadapt; Site_1: 5' EcoRI;
                     Site_2: 3' XhoI; Roots were harvested from plants grown
                     under the following deficiencies/stresses: 10 mM Al, Zn,
                     P, K, Fe,N. mRNA was isolated from indivdual treatments.
                     Proportional aliquots of mRNA of each treatment were mixed
                     and used for library construction."
ORIGIN

Alignment Scores:
Pred. No.:              3.24e-12      Length:        577
Score:                  191.50        Matches:       43
Percent Similarity:     50.88%        Conservative:  44
Best Local Similarity:  25.15%        Mismatches:    71
Query Match:            6.36%         Indels:        13
DB:                     10            Gaps:          2

US-09-913-878A-2_COPY_710_1282 (1-573) x BF096404 (1-577)

Qy        165 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAsp 184
               |||         :::::::::::||||||  |||:::   :::   :::|||:::
Db         51 ACAGCGGCTATAATGTTGAGTGCAGGGTTTAAACCTCAAAGTGAACCTCATTTGAGAGGG 110

Qy        185 IleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArg 204
               :::      ::::::      :::            |||::::::|||      :::
Db        111 ATGTTGGCTAGCATTAGAGCTGCTCAGCTTGGCGACCTCAGGAATAAGACAAGGATGTTT 170

Qy        205 ValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsn 224
               |||            ::::::         |||      |||   ||||||:::
Db        171 GTTACTTCAGGAAGGTGGTTGATGGGCTGTTTGGATGAATTAGGTGAACTTGAGCAAGGC 230

Qy        225 GluValHisValGlyPheSerSer------------------------------------ 232
               :::      :::   ||||||
Db        231 CAATGCTTTATTCAAGTGTCAAGCCCTTCTTTGGAGACCTGTTTTGTTAAGCATGGTCCA 290

Qy        233 LysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAla 252
               :::|||      :::::::      ::::::               |||::::::|||
Db        291 GAGTTTTCTGAAATCAAGAAAAATCTTCAAGTAGTAAAGGGCCTT---GTTGTAATTGCA 347

Qy        253 ArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGlu 272
               ::::::|||         |||   |||::::::   :::   ||||||      |||
Db        348 AAGAACCCGTGTCTTCATCCCGGGGATGTGAGGATTCTGGAGGCTGTAGATGTTCCTGGT 407

Qy        273 LeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLys 292
               ||||||   |||   |||   ::::::|||         |||||||||   |||   :::
Db        408 TTACACCATCTCTATGATTGTCTGGTCTTCCCTCAGAAAGGGGATAGGCCACATTCAAAT 467

Qy        293 LysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIle 312
               :::   ||||||   |||   |||||||||:::   :::|||   ||||||      :::
Db        468 GAAGCATCAGGGAGTGACCTTGATGGTGATCTCTACTTTGTGACTTGGGATGAAAATCTC 527

Qy        313 ValAspGlyPheValAsnAlaGluMetProLeu 323
               :::         :::   ||||||:::
Db        528 ATTCCACCCAGTAAGAAAAGCTGGATGCCAATG 560
```

```
RESULT 4
CG445102/c
LOCUS       CG445102                 954 bp    DNA     linear   GSS 17-SEP-2003
DEFINITION  OG8AH57TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0850J18,
            genomic survey sequence.
ACCESSION   CG445102
VERSION     CG445102.1  GI:34827337
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 954)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..954
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0850J18"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

Alignment Scores:
Pred. No.:              5.37e-11      Length:        954
Score:                  184.50        Matches:       61
Percent Similarity:     41.51%        Conservative:  49
Best Local Similarity:  23.02%        Mismatches:    91
Query Match:            6.13%         Indels:        65
DB:                     29            Gaps:          8

US-09-913-878A-2_COPY_710_1282 (1-573) x CG445102 (1-954)

Qy        162 GlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyr 181
                |||      |||         ::::::   :::||||||   |||      :::   :::
Db        818 CAAGGAAGCACCGCAGGGCTGATGCTGGGCGCTGGCTTCGCTCCTGCGACTGAGCCACAC 759

Qy        182 LeuGlnAspIleAla-TrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLe 201
               |||:::   :::   |||         |||:::|||               :::|||
Db        758 CTGAGAGCGATGCTACTGGCCATCTGCACCTCGCAGATGCAGGGCCTTCTCAACAAAACG 699

Qy        201 uAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLe 221
                 |||   |||   :::      ::::::         |||   :::|||:::||
Db        698 TGG-ATCTTTGTGCCCAAGGAAGGTGGCTGATGGGCTGTCTCGACGAGTTTGGGATCCT 640

Qy        221 uGluGluAsnGluVal--------------------HisValGly------------- 229
               ||||:::   :::                        ||||||
Db        639 TGAGCAGGGGCAGTGCTTTCTCCGGGTCTCGACGCCATCAGTCGGCAGCCGCTGCGTGAA 580

Qy        230 ------------PheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeuSe 245
                          |||   |||:::::::::   :::   ::::::      :::
Db        579 CCGTTCTCCGATATTTCCCTCGGAGTATAAGGGCGATGCCGATGCACAGGTGATTACTGG 520

Qy        245 rAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgVa 265
                     |||::::::|||::::::|||            |||   |||::::::   ::
Db        519 T------ACGGTTGTGATGGCGAAGAACCCATGCCTTCATCCAGGGGACGTCCGTATCCT 466
```

```
Qy        265 lArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLy 285
              :   ||||||        |||:::|||:::    |||   |||       :::|||       ||
Db        465 TGAAGCTGTGGATGTCCCTGATCTGTATCACCTTGTCGACTGCTTCGTCTTCCCCAAGAA 406

Qy        285 sGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTr 305
              ||||:::   |||   |||    :::   |||||   |||   ||||||||:::    ::
Db        405 AGGTGAGAGGCCACACGCCGACGAAGCTTCCGGGAGCGACCTCGACGGCGACGTCTACTT 346

Qy        305 pValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGlu-- 324
              :|||   ||||||        :::|||               :::      |||::::::
Db        345 CGTGACGTGGGACGAGAGCCTTGTACCTCCCGGGAGGAAGAGCTGCACCCCGATGGACTA 286

Qy        325 ----ProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAl 343
                   |||        ::::::   |||    :::|||
Db        285 CTCCCCGGCTAAAACAAAGCAACTTCCACGCGAC------------------------- 252

Qy        343 aSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSe 363
                                                    :::        :::
Db        251 ------------------------------------GTACACCAACAAGTAAGTACACA 229

Qy        363 rPheHisPheAlaLeuGlnProAsnPhe---LeuGlyMetCysThrAsn---------- 378
                 :::||||||||         :::|||    |||       |||    |||
Db        228 ATACCACTTTGCT------TCTTCCTTTGGTTTATCTTGCTGCCATAATACATCAAGTTA 175

Qy        379 --------------------------------------------TyrLysGluArgLe 383
                                                          |||    |||    ::
Db        174 ATGGATTTCTCGCGTGCTCGTGTTGTAACAGGATACCATTGATTTCTACTTGGAGAGCAT 115

Qy        383 uCysTyrIleAsn 387
              :   |||   |||
Db        114 GGTATACGACAAC 102
```

```
RESULT 5
CA257764
LOCUS       CA257764                 641 bp    mRNA    linear   EST 26-SEP-2003
DEFINITION  SCCCFL1100F01.g FL1 Saccharum officinarum cDNA clone SCCCFL1100F01
            5', mRNA sequence.
ACCESSION   CA257764
VERSION     CA257764.1  GI:35945014
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 100  row: F  column: 01
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1..641
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCCCFL1100F01"
                     /lab_host="DH10B"
                     /clone_lib="FL1"
                     /note="Organ: Inflorescence at begining of development
                     (1cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI;
                     An unidirectional cDNA library generated from
                     [Inflorescence at begining  of development (1cm-long)].
                     cDNA was prepared from polyA+ mRNA using  SuperScript
                     Plasmid System Kit (Invitrogen). The double-strand cDNAs
                     were  fractionated in a sepharose CL-2B 40cm-columns and
                     fragments sizing  between 0.8 and 1.5 Kb were
                     directionally cloned into the vector. Details  of each
                     source of RNA and library construction can be obtained at
                     http://sucest.lad.ic.unicamp.br/public"
ORIGIN

Alignment Scores:
Pred. No.:                3.66e-11      Length:         641
Score:                    183.50        Matches:        44
Percent Similarity:       50.30%        Conservative:   39
Best Local Similarity:    26.67%        Mismatches:     67
Query Match:              6.10%         Indels:         15
DB:                       14            Gaps:           3

US-09-913-878A-2_COPY_710_1282 (1-573) x CA257764 (1-641)
```

```
Qy        162 GlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyr 181
              |||      |||        ::::::::::::|||      |||      :::   :::
Db         25 CAAGGAAACACTGCAGCACTGATGTTGAGTGCTGGCATTTCACCTGGAACTGAACCACAC 84

Qy        182 LeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeu 201
              |||:::   :::      ::::::      :::          |||      |||
Db         85 CTGAAAGCAATGCTGTTAGCTATAAGGTCCTCACAGCTACTAGGCCTTTTGGAGAAAACA 144

Qy        202 AsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeu 221
                 |||   |||     :::      ::::::             |||      |||:::|||
Db        145 AGGATTTTTGTGCCCAAGGGGAGGTGGTTGATGGGCTGCCTTGATGAACTTGGGATCCTT 204

Qy        222 GluGluAsnGluValHisValGlyPheSerSer-------------------------- 232
              |||:::   :::      :::      ||||||
Db        205 GAGCAAGGACAGTGCTTTATCCGGGCGTCATCTCCATCACTCAATAATTGTCTAGTAAAG 264

Qy        233 ---------LysPhe---ArgAspGluGluGluSerPheThrLeuLeuSerAspCysAsp 248
                       :::|||   :::      :::   :::   |||::::::
Db        265 CATGGACCAAGATTTTCCAAAGCAAACAAAAATGCAGAGACCATTGTGGGT------ACT 318

Qy        249 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 268
              ::::::::|||::::::|||               |||   |||::::::   :::   ||||||
Db        319 ATTGTAATGGCAAAGAATCCATGTCTTCATCCAGGGGATGTCCGGATTCTTGAAGCTGTT 378

Qy        269 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 288
                    ||||||||||||      |||   |||      ::::::|||      ||||||
Db        379 GATGTGCCTGAACTGCATCACCTTGTTGATTGCTTGGTCTTCCCTAAGAAAGGTGATAGG 438

Qy        289 ProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrp 308
              |||   |||   :::   |||||   |||   ||||||||:::   :::|||   |||
Db        439 CCACATGCCAATGAAGCATCTGGGAGTGATCTTGATGGGGATCTCTACTTTGTGACATGG 498

Qy        309 AspProGluIleVal 313
              |||      ::::::
Db        499 GATGAAAACCTTATA 513
```

```
RESULT 6
BI246074
LOCUS       BI246074                 582 bp    mRNA    linear   EST 17-JUL-2001
DEFINITION  IP1_68_D10.b1_A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BI246074
VERSION     BI246074.1  GI:14824093
KEYWORDS    EST.
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and
            Pratt,L.H.
  TITLE     An EST database from Sorghum: developing preanthesis pannicles
```

```
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for high quality sequence is
            20. Three-prime sequences, which are obtained with PolyTMix or T7
            sequencing primer, are presented as the reverse complement.
            Seq primer: JEN REV
            High quality sequence stop: 560
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Sorghum bicolor"
                     /mol_type="mRNA"
                     /cultivar="BTx623"
                     /db_xref="taxon:4558"
                     /clone_lib="Immature pannicle 1 (IP1)"
                     /note="Organ: Developing preanthesis pannicles; Vector:
                     pBluescript II SK(-) from Lambda Zap II; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
ORIGIN

Alignment Scores:
Pred. No.:              2.95e-10      Length:        582
Score:                  175.50        Matches:       42
Percent Similarity:     49.69%        Conservative:  38
Best Local Similarity:  26.09%        Mismatches:    60
Query Match:            5.83%         Indels:        21
DB:                     12            Gaps:          3

US-09-913-878A-2_COPY_710_1282 (1-573) x BI246074 (1-582)

Qy        169 LeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAsp 188
              :::::::::::|||         |||      :::   :::||||:::   :::
Db         17 ATGTTGAGTGCTGGCATTTCACCTGGAACTGAGCCACACCTGAAAGCAATGCTGTTAGCT 76

Qy        189 LeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSer 208
              ::::::      :::         |||      |||      |||   |||   :::
Db         77 ATAAGGTCCTCACAGCTACTGGGTCTTTTGGAGAAGACAAGGATTTTTGTGCCGAAGGGG 136

Qy        209 AlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisVal 228
              ::::::                  |||      |||:::||||||:::   :::      :::
Db        137 AGGTGGTTGATGGGCTGCCTTGATGAACTTGGGATCCTTGAGCAAGGACAGTGCTTTATC 196

Qy        229 ------------------------------------------------GlyPheSerSer 232
                                                              ||||||||||
Db        197 CGGGCGTCATCTCCATCACTCAATAATTGTCTGGTAAAGCATGGACCAAGATTTTCCTCA 256

Qy        233 LysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAla 252
                    ::::::   |||         |||::::::          :::::::::|||
Db        257 GCAAACAAAAATGCAGAG---------ACCATTGTGGGT------ACTATCGTAATGGCA 301

Qy        253 ArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGlu 272
              ::::::|||         |||   |||::::::   :::   ||||||         ||||||
Db        302 AAGAATCCATGCCTTCATCCAGGGGATGTCCGGATCCTTGAAGCTGTTGATGTGCCTGAA 361

Qy        273 LeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLys 292
               ||||||   |||   |||   ::::::|||      ||||||:::   |||   |||
Db        362 CTGCATCACCTTGTTGATTGCTTGGTCTTCCCCAAGAAAGGTGAGAGGCCACATGCCAAT 421

Qy        293 LysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIle 312
              :::   ||||||   |||   |||||||||:::   :::|||   ||||||         :::
Db        422 GAAGCATCTGGGAGTGATCTTGATGGGGATCTCTACTTTGTGACATGGGATGAAAACCTT 481

Qy        313 Val 313
              :::
Db        482 ATA 484

RESULT 7
BU965311
LOCUS       BU965311                 622 bp    mRNA    linear   EST 21-OCT-2002
DEFINITION  sat08g07.y1 Gm-c1036 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1036-13357 5' similar to TR:Q9ZR58 Q9ZR58 RNA-DIRECTED RNA
            POLYMERASE ;, mRNA sequence.
ACCESSION   BU965311
VERSION     BU965311.1  GI:24206058
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact: ccu@resgen.com web site:
            www.resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 445.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1036-13357"
                     /tissue_type="somatic embryos cultured on MSD 20"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1036"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
                     cDNA library was constructed from mRNA isolated from
                     somatic  embryos (age ranging from 2 months to 9 months)
                     cultured on MSD 20. The library was prepared using the
                     Life Technologies pSuperScript cDNA library construction
                     kit. Complementary DNA was synthesized from mRNA using a
                     poly (dT) sequence with a NotI restrictions site. SalI
                     linkers adapters were ligated to the blunt-ended cDNA
                     fragments followed by NotI digestion. The cDNA fragments
                     were directionally cloned into the NotI-SalI restriction
                     site of the pSPORT1 vector. The ligated cDNA fragments
                     were transformed into E.coli ElectroMax DH10B host cells.
                     This library was constructed in the laboratory of Dr. Lila
                     Vodkin by Anu Khanna at the University of Illinois at
                     Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"
ORIGIN

Alignment Scores:
Pred. No.:              6.67e-10      Length:        622
Score:                  173.00        Matches:       44
Percent Similarity:     50.00%        Conservative:  34
Best Local Similarity:  28.21%        Mismatches:    70
Query Match:            5.75%         Indels:        8
DB:                     13            Gaps:          2
```

```
US-09-913-878A-2_COPY_710_1282 (1-573) x BU965311 (1-622)

Qy        166 LeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIle 185
Db        149 CTGAAGGAGATGCTCATTTGTGGATACAAGCCTAATGAAGAACCATTCCTTTCAATGATG 208

Qy        186 AlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgVal 205
Db        209 CTTCAAACATTTAGGGCATCAAAACTTTTGGAATTGCGACTTAAATCTAGGATCTTTATT 268

Qy        206 GlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGlu 225
Db        269 CCAAAAGGAAGAGCAATGATGGGATGTCTAGATGAAACTAGAACCCTAGAATATGGTCAA 328

Qy        226 ValHisValGlyPheSerSer------------LysPheArgAspGluGluGluSerPhe 241
Db        329 GTATTTGTTCAGTTCTCTAACAATAGGCTGCAGAATCTATCTGATGATTTTTTTCATAT 388

Qy        242 ThrLeu------------LeuSerAspCysAspValLeuValAlaArgSerProAlaHis 257
Db        389 GATTTGCCAAAGAATTATATGGTTAAAGGTAAGGTAGTTGTAGCAAAAAACCCCTGCTTG 448

Qy        258 PheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLys 277
Db        449 CACCCAGGTGATGTGCGTGTTTTACAAGCTGTGGATGTGCCAGATTTGTACCACATGGTG 508

Qy        278 AspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGly 297
Db        509 GACTGTGTTGTTTTCCCTCAAAAAGGACCAAGACCTCATCCAAATGAGTGTTCGGGAAGT 568

Qy        298 AspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleVal 313
Db        569 GATCTGGATGGAGATATCTACTTTGTTTGTTGGGACCATGAAATGATT 616

RESULT 8
BF275078
LOCUS       BF275078                761 bp    mRNA    linear   EST 07-MAR-2001
DEFINITION  GA__Eb0023B06f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Eb0023B06f, mRNA sequence.
ACCESSION   BF275078
VERSION     BF275078.1  GI:11206148
KEYWORDS    EST.
SOURCE      Gossypium arboreum
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 761)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J.,
            Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 751.
FEATURES             Location/Qualifiers
     source          1..761
                     /organism="Gossypium arboreum"
                     /mol_type="mRNA"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Eb0023B06f"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Alignment Scores:
Pred. No.:              1.24e-09      Length:         761
Score:                  172.00        Matches:        66
Percent Similarity:     40.14%        Conservative:   46
Best Local Similarity:  23.66%        Mismatches:     102
Query Match:            5.71%         Indels:         65
DB:                     10            Gaps:           9

US-09-913-878A-2_COPY_710_1282 (1-573) x BF275078 (1-761)

Qy        232 SerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 251
Db         36 TCTAGGTTTGCTGAAACAAAGAAAAATTGGGAAGTAATTAAGGGGCTG---GTTGTCATA 92

Qy        252 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 271
Db         93 GCTAAGAATCCTTGTCTTCACCCTGGAGATATAAGAATTCTAGAAGCAGTTGATGCCCCT 152

Qy        272 GluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAla 291
Db        153 GGTTTACACCATTTGTATGACTGTCTGGTCTTCCCTCAAAAGGGTGAGAGACCCCATACG 212

Qy        292 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 311
Db        213 AATGAAGCTTCTGGAAGTGATCTTGATGGGGACCTTTATTTTGTCACATGGAAGAGGCT 272

Qy        312 IleValAspGlyPheValAsnAlaGluMetProLeu------GluProAspLeuSerArg 329
Db        273 CTTATCCCCCCTAGCAAGAAGAGCTCACAACCCATGCAGTATGACCCTGATGAGCCTCGA 332

Qy        330 TyrLeuLysLysAspLysThr------------------ThrPheLysGlnLeuMetAla 343
Db        333 GAGTTGAATCGCCCAGTCACTCATAAGGACATAATTGAGTTTTTTCAAAAAACATGGTG 392

Qy        344 SerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSer 363
Db        393 AATGAGCACCTGGGAAGTATCTGCAATGCACATGTGGTTCATTCTGACCTTAGTGAA--- 449

Qy        364 PheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeu 383
Db        449 ------------------------------------------------------------ 449

Qy        384 CysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGly 403
Db        450 ------------CATGGTGCTTCAGATGAGAAATGTATACATCTCGCAGAGTTAGCGGCT 497

Qy        404 AsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeu 423
Db        498 ATAGCTGTTGATTTTCCTAAAACCGGGAAGATTGTGTCGATGCCT------GCTCAATTA 551

Qy        424 ArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMet--------------- 438
Db        552 AAACCAAAACTT------------------TACCCAGATTTATGGGTAAAGAGGAGTTC 593

Qy        439 -------TyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTy 456
Db        594 CAGTCATTACAAGTCAAACAAAATTCTGGGAAGA------------------------- 627

Qy        456 rLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAl 476
Db        628 ----CTATATCGTTATATCAGAGATGCTTATGACCAAGATGTCTCTGAATCTTCTGAGCT 683

Qy        476 aMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeu 494
Db        684 CAATTTTGGTGCTAGTGACATC---------------ATTTATGATGCCGATCTT 723

RESULT 9
CG212111
LOCUS       CG212111                934 bp    DNA     linear   GSS 22-AUG-2003
```

```
DEFINITION  OG0FJ51TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0705J06,
            genomic survey sequence.
ACCESSION   CG212111
VERSION     CG212111.1  GI:34111941
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 934)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OG0FJ51TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..934
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0705J06"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

Alignment Scores:
Pred. No.:              2.31e-09       Length:         934
Score:                  171.00         Matches:        67
Percent Similarity:     37.84%         Conservative:   45
Best Local Similarity:  22.64%         Mismatches:     84
Query Match:            5.68%          Indels:         100
DB:                     29             Gaps:           12

US-09-913-878A-2_COPY_710_1282 (1-573) x CG212111 (1-934)

Qy        216 AspPheTrpGlyValLeuGluGluAsnGluVal---------------------HisVal 228
              |||   :::|||:::|||||:::   :::                         |||
Db         27 GACGAGTTTGGGATCCTTGAGCAGGGGCAGTGCTTTCTCCGGGTCTCGACGCCATCAGTC 86

Qy        229 Gly---------------------------PheSerSerLysPheArgAspGluGluGlu 239
              |||                           |||   |||:::::::::   :::   :::
Db         87 GGCAGCCGCTGCGTGAACCGTTCTCCGATATTTCCCTCGGAGTATAAGGGCGATGCCGAT 146

Qy        240 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 259
              :::      :::                  |||::::::|||::::::|||      |||
Db        147 GCACAGGTGATTACTGGT------ACGGTTGTGATGGCGAAGAACCCATGCCTTCATCCA 200

Qy        260 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 279
                 |||::::::   :::   ||||||      |||:::|||:::   |||   |||
Db        201 GGGGACGTCCGTATCCTTGAAGCTGTGGATGTCCCTGATCTGTATCACCTTGTCGACTGC 260

Qy        280 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyr 299
                 :::|||      ||||||:::   |||   |||   :::   ||||||   |||
Db        261 TTCGTCTTCCCCAAGAAAGGTGAGAGGCCACACGCCGACGAAGCTTCCGGGAGCGACCTC 320

Qy        300 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAla 319
              |||||||||:::   :::|||   ||||||      :::|||               :::
Db        321 GACGGCGACGTCTACTTCGTGACGTGGGACGAGAGCCTTGTACCTCCCGGGAGGAAGAGC 380

Qy        320 GluMetProLeuGlu------ProAspLeuSerArgTyrLeuLysLysAspLysThrThr 337
                 |||::::::      |||      ::::::   |||   :::|||
Db        381 TGCACCCCGATGGACTACTCCCCGGCTAAAACAAAGCAACTTCCACGCGAC--------- 431

Qy        338 PheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyr 357
                                                                        :::
Db        432 ---------------------------------------------------GTACAC 437

Qy        358 AspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPhe---LeuGlyMetCys 376
                 :::         :::|||||||||         :::|||   |||      |||
Db        438 CAACAAGTAAGTACACAATACCACTTTGCT------TCTTCCTTTGGTTTATCTTGCTGC 491

Qy        377 ThrAsn------------------------------------------------------ 378
              |||
Db        492 CATAATACATCAAGTTAATGGATTTCTCGCGTGCTCGTGTTGTAACAGGATACCATTGAT 551

Qy        379 ---TyrLysGluArgLeuCysTyrIleAsn------------------------------ 387
                 |||   |||   :::   |||   |||
Db        552 TTCTACTTGGAGAGCATGGTATACGACAACCTTGGTCGAATATGCAACGCCCATGCTGCT 611

Qy        388 ------------------AsnSerValSerAsnLysProAlaIleIleLeuSerSerLeu 401
                                :::         :::         |||   |||:::   |||
Db        612 CATGCTGATCGGAGCGACGACGGAGCGATGGATCCCAAGTGCATCGAGTTGGCTCGGCTC 671

Qy        402 ValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAla 421
                             ||||||   :::|||   |||   :::
Db        672 GCATCTATCGCCGTGGACTCCGCCAAGACCGGGGAAATTGTAAGA--------------- 716

Qy        422 GlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSer---------LeuProAspProMet 438
                                :::   |||||||||         ||||||   |||
Db        717 -----------------ATGCCACCAGCCCTTTCGCCAAAAGAGTACCCCGACTTCATG 758

Qy        439 ---------------------TyrLysSerAspSerTrpLeuGlyArg 447
                                   |||:::|||:::   |||||||||
Db        759 GGGAAGGAGGACGCCATCTCTTACAGATCCGAAAAGATCCTTGGGAGG 806

RESULT 10
CA223511
LOCUS       CA223511               648 bp    mRNA    linear   EST 25-SEP-2003
DEFINITION  SCJFFL1C03G07.g FL1 Saccharum officinarum cDNA clone SCJFFL1C03G07
            5', mRNA sequence.
ACCESSION   CA223511
VERSION     CA223511.1  GI:35279914
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: C03  row: G  column: 07
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1..648
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCJFFL1C03G07"
                     /lab_host="DH10B"
                     /clone_lib="FL1"
                     /note="Organ: Inflorescence at begining of development
                     (1cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI;
```

```
                An unidirectional cDNA library generated from
                [Inflorescence at begining  of development (1cm-long)].
                cDNA was prepared from polyA+ mRNA using  SuperScript
                Plasmid System Kit (Invitrogen). The double-strand cDNAs
                were  fractionated in a sepharose CL-2B 40cm-columns and
                fragments sizing  between 0.8 and 1.5 Kb were
                directionally cloned into the vector. Details  of each
                source of RNA and library construction can be obtained at
                http://sucest.lad.ic.unicamp.br/public"
ORIGIN

Alignment Scores:
Pred. No.:              1.91e-09      Length:        648
Score:                  169.50        Matches:       59
Percent Similarity:     41.91%        Conservative:  42
Best Local Similarity:  24.48%        Mismatches:    111
Query Match:            5.63%         Indels:        29
DB:                     14            Gaps:          6

US-09-913-878A-2_COPY_710_1282 (1-573) x CA223511 (1-648)

Qy         85 LeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLysMetArg 104
               |||||    :::|||:::   :::|||
Db         11 CTGAATCGGAAGTTGATTACTCTTCTCTCA---------------------------ACA 43

Qy        105 GlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHis 124
                  :::||||||    :::                |||::::::           |||
Db         44 CTTGGGGTTGGCGATAGTGTCTTTGAACTAAAGCAGAAGGAAGTCGTAAGGCAG------ 97

Qy        125 AlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAla 144
                 |||||||||   |||                                       :::|||
Db         98 ---TTGAACAGAATGGTAACC---------------------------GAACCACAGGCT 127

Qy        145 ThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluGlu 164
                 |||    :::          :::|||
Db        128 GCTCGTGAAGCGGTTGAACTTATGCCCATG---------------GGAGAAGTTACCAAT 172

Qy        165 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAsp 184
                :::          |||::::::   |||:::   |||       :::   ||||||
Db        173 GTAGTTAAAGAATTGTTATCATGTGGCTACCAGCCTGATCATGAGCCGTATCTTTCCATG 232

Qy        185 IleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArg 204
              :::          :::         |||         ||||||:::|||         |||
Db        233 CTGCTACAAACTTTTAGAGCATCCAAGCTTCTAGAATTGAAAACAAAGTCAAGGATATTC 292

Qy        205 ValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsn 224
              :::   :::         :::               |||         |||:::
Db        293 ATCCCACAAGGGCGAGCAATGATGGGTTGCCTGGATGAAACCCGCACACTAAAGTATGGC 352

Qy        225 GluValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeu 244
              :::|||         :::|||            |||         |||      :::
Db        353 CAGGTATTCATCCAA---GCTTCTTACTGTGCAGATGACCATCGCAAGTTCGTTGTAACT 409

Qy        245 SerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArg 264
                                |||:::||||||::::::|||      |||   ||||||:::
Db        410 GGA------AAAGTAGTTGTTGCCAAAAATCCTTGTCTCCACCCTGGTGACATACGGGTT 463

Qy        265 ValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThr 284
              ::::::||||||      |||   ||||||   |||   |||   ::::::|||
Db        464 CTCCAAGCTGTTGATATTCCTGCTCTGCACCACTTGTTTGACTGTGTTGTCTTTCCACAG 523

Qy        285 LysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAla 304
              :::|||         |||         :::   ||||||   |||   |||||||||:::
Db        524 CAGGGACCAAGGCCGCACCCTAATGAGTGTTCAGGGAGTGATCTTGATGGGGACATATAT 583

Qy        305 TrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGlu 324
              :::|||   ||||||      ::::::         :::            |||
Db        584 TTTGTTTCTTGGGATTCACATCTTATTCCAAGTCGTTTGGTGGATCCTATGGACTATTCT 643

Qy        325 Pro 325
              |||
Db        644 CCA 646
```

```
RESULT 11
CF441078
LOCUS       CF441078                 741 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  EST677423 normalized cDNA library of onion Allium cepa cDNA clone
            ACAGB56, mRNA sequence.
ACCESSION   CF441078
VERSION     CF441078.1  GI:34463768
KEYWORDS    EST.
SOURCE      Allium cepa (onion)
  ORGANISM  Allium cepa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
            Allium.
REFERENCE   1  (bases 1 to 741)
  AUTHORS   Havey,M.J., Cheung,F., Van Aken,S., Utterback,T. and Town,C.D.
  TITLE     Expressed Sequence Tags from a normalized library of mixed onion
            tissues (Allium cepa)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Havey MJ
            Department of Horticulture
            USDA-ARS and University of Wisconsin
            1575 Linden Drive, Madison, WI 53706, USA
            Tel: 608-262-1830
            Fax: 608-262-4743
            Email: mjhavey@facstaff.wisc.edu
            TIGR sequence name ACAGB56TR. For more information:
            http://haveylab.hort.wisc.edu
            Seq primer: CAG GAA ACA GCT ATG ACC.
FEATURES             Location/Qualifiers
     source          1..741
                     /organism="Allium cepa"
                     /mol_type="mRNA"
                     /cultivar="Red Creole(bulbs), unknown(callus), Ebano &
                     Texas Legend(roots)"
                     /db_xref="taxon:4679"
                     /clone="ACAGB56"
                     /tissue_type="Callus, roots, and young bulbs"
                     /clone_lib="normalized cDNA library of onion"
                     /note="Vector: pCMVSport6.1-ccdb (Invitrogen); Site_1:
                     EcoRV (5'); Site_2: NotI (3'); Equal molar amounts of mRNA
                     from callus, roots, and young bulbs were combined to
                     synthesize the library. Normalization to enrich for
                     low-copy transcripts was performed by proprietary
                     techniques of Invitrogen."
ORIGIN

Alignment Scores:
Pred. No.:              2.39e-09      Length:        741
Score:                  169.50        Matches:       64
Percent Similarity:     40.94%        Conservative:  49
Best Local Similarity:  23.19%        Mismatches:    106
Query Match:            5.63%         Indels:        57
DB:                     14            Gaps:          8

US-09-913-878A-2_COPY_710_1282 (1-573) x CF441078 (1-741)

Qy        197 LeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAsp 216
               |||      |||      |||   |||   :::      ::::::            |||
Db          3 CTGGCGATAAAAACCAGGATATTTGTGAATAAAGGGAGGTGGTTGATGGGATGTTTGGAT 62

Qy        217 PheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSer------------ 232
                       |||:::||||||:::      :::      :::         ||||||
Db         63 GAACTTGGGCTGTTAGAGCAAGGGCAGTGCTTTATTCAGGTTTCCAGTCCTTGTGTTGAG 122

Qy        233 ------------------------LysPheArgAspGluGluGluSerPheThrLeuLeu 244
                                      |||:::         :::|||            ::::::
Db        123 AATGCCTTTTTGAAGTATGGGGAGTCTTTTAAGAGTAGGAAGGAGAGGAAAGAGGTGATA 182

Qy        245 SerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArg 264
                       ::::::   |||::::::|||            |||   |||::::::
Db        183 AAGGGT---ACAATAGTTTTCGCCAAGAATCCATGTTTGCACCCTGGTGATGTTAGGGTA 239

Qy        265 ValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThr 284
```

```
                :::    |||:::        ||||||||        |||   |||   ::::::|||
Db        240 CTTGAAGCGATTGATGTGCCGGAATTGAAGCATCTTGTGGACTGTTTGGTTTTTCCTCAG 299

Qy        285 LysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAla 304
              ||||||||     |||    :::   :::    ||||||     |||   |||||||||:::
Db        300 AAAGGAGATAGACCGCATTCAAATGAGGCTAGTGGAAGCGACCTTGATGGCGATTTGTAT 359

Qy        305 TrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGlu 324
              :::|||  ||||||       ::::::            ::::::      |||::::::
Db        360 TTTGTTACTTGGGATGAGAGGTTGATTCCACCGGGTAAAAGCAGTTGGCCTCCGATGGAT 419

Qy        325 ProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSer 344

Db        419 ------------------------------------------------------------ 419

Qy        345 HisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPhe 364
                   |||   :::   ||||||::::::                 |||
Db        420 ---TATACTGCAGCTGAGGCGAAGGATGAGCCA---------------CAACCCATTGTT 461

Qy        365 HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCys 384
              |||       :::        |||:::          ::::::     |||   |||
Db        462 CATCAAGATATTTTAACCTTCTTCATCAAGAACGTATCGAGT------GAGAACCTAGGA 515

Qy        385 TyrIleAsnAsnSer---------------------------------ValSerAsnLys 393
                 |||:::|||:::                                       ::::::
Db        516 AGGATCAGCAATGCTCATGTAGTTCATGCAGACCAAAGCGACTATGGTGCCATGGACGAA 575

Qy        394 ProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGly--- 412
                    |||   |||:::  |||           ||||||       |||   |||
Db        576 AAATGCATCAAACTTGCAGAGCTTGCGGCTGAAGCAGTCGACTTCCCTAAAACCGGTAAA 635

Qy        413 ---IleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAla 431
                 :::                 :::|||    :::     :::|||
Db        636 ATAGTAAAGATGCCTCCGTCCCTGAAGCCCAAACTGTACCCTGACTTCATGGGCAAGGAC 695

Qy        432 LeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArg 447
                 ||||||          |||||||||:::      |||||||||
Db        696 AGTTCCCTT---------ACCTACAAATCAGAGAAAATACTGGGCAGA 734
```

RESULT 12
BQ827639

```
LOCUS       BQ827639              541 bp    mRNA    linear   EST 07-AUG-2002
DEFINITION  gd58f03.y1 Moss EST library PPAS Physcomitrella patens cDNA clone
            PEP_SOURCE_ID:PPAS041205 5' similar to TR:Q9SG02 Q9SG02 PUTATIVE
            RNA-DIRECTED RNA POLYMERASE. ;, mRNA sequence.
ACCESSION   BQ827639
VERSION     BQ827639.1  GI:22133265
KEYWORDS    EST.
SOURCE      Physcomitrella patens
  ORGANISM  Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Quatrano,R., Bashiardes,S., Cove,D., Cuming,A., Knight,C.,
            Clifton,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wylie,T.,
            Underwood,K., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Steptoe,M., Gibbons,M., Harvey,N., Ritter,E.,
            Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
  TITLE     Leeds/Wash U Moss EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Libraries were constructed by Dr. Stavros Bashiardes as part of the
            Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
            Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1..541
                     /organism="Physcomitrella patens"
                     /mol_type="mRNA"
                     /db_xref="taxon:3218"
                     /clone="PEP_SOURCE_ID:PPAS041205"
                     /dev_stage="protonemata, 7day old ABA-treated"
                     /lab_host="E.coli DH10b"
                     /clone_lib="Moss EST library PPAS"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     EcoRI; 7-day-old protonemal tissue was incubated 10-4M ABA
                     for 16 hours before RNA isolation. cDNA was synthesised
                     from Amersham's cDNA Sythesis Plus kit and ligated with
                     EcoRI-NotI linkers for cloning in the EcoRI site of Lampda
                     ZapII (Stratagene). After packaging, the library was
                     propogated in E. coli XL-I Blue cells and amplified. The
                     library was excised by mass excision in SOLR cells and
                     ampicillan resistant transformants selected.
                     Approximately, 1,000,000 colonies were grown and
                     recovered. The double stranded plasmid library was
                     recovered and used to transform DH10b cells by
                     electroporation. Clones corresponding to abundant
                     transcripts were identified by colony hybridization using
                     a cDNA probe derived from untreated protonematal tissue,
                     and eliminated from the library, by rearraying."
ORIGIN
```

```
Alignment Scores:
Pred. No.:              2.15e-09        Length:          541
Score:                  168.00          Matches:         42
Percent Similarity:     44.79%          Conservative:    31
Best Local Similarity:  25.77%          Mismatches:      64
Query Match:            5.58%           Indels:          26
DB:                     13              Gaps:            3
```

US-09-913-878A-2_COPY_710_1282 (1-573) x BQ827639 (1-541)

```
Qy        169 LeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAsp 188
              ::::::               ||||||   |||   |||:::   :::|||   ::::::
Db         68 ATGCTTCGCGGAGGATTTCTCCCTTTGAAGGAGCCTCATCTCTATGAATTACTGGTCTCT 127

Qy        189 LeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSer 208
              ::::::               :::   |||   ||||||       |||   |||
Db        128 ATCAAGTTCAGCCTGCTGGAGGATTTAGTTTCAAAGGCTAGAATATTTGTACCTAATGGA 187

Qy        209 AlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGlu------Val 226
                 ::::::            |||      |||:::|||          :::      :::
Db        188 CGTTGGGTCATTGGGTGTATGGACGAGACTGGGTTGTTGAATTACGGGCAATGTTTCATT 247

Qy        227 HisVal------------------------------------------------GlyPhe 230
              ||||||                                                   |||
Db        248 CATGTATCTGGCCCTGTAAGCAAGGCCTGCCTTGGAGGAGATGGAAATAATCTGGGACAT 307

Qy        231 SerSerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeu 250
              |||      :::                             :::            |||:::
Db        308 AGTGGACTCCTGCAG------------------------GTGATCACGGGAAAGGTTATT 343

Qy        251 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 270
              :::   ::::::|||        |||   ||||||:::   :::   ||||||
Db        344 ATGGTAAAAAATCCATGCTTGCATCCTGGAGATATCCGCATACTGGAAGCAGTGGATCTG 403

Qy        271 ProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeu 290
              |||||||||||||||:::   |||   ::::::                |||      |||
Db        404 CCGGAATTGCACAGCATGGTCGACTGCTTGGTTTTACCACAGAATGGACCAAGACCGCAT 463

Qy        291 AlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspPro 310
                 ::::::||||||   |||   |||||||||:::   :::   ||||||||||||
Db        464 CCAAATGAGGTCTCTGGCAGTGATCTGGATGGGGATGTATACTTCACTTGTTGGACCCG 523

Qy        311 GluIleVal 313
                 ::::::
Db        524 AATTTGATT 532
```

```
RESULT 13
CD928156
LOCUS       CD928156                 677 bp    mRNA    linear   EST 15-JUL-2003
DEFINITION  GR45.104C10F010316 GR45 Triticum aestivum cDNA clone GR45104C10,
            mRNA sequence.
ACCESSION   CD928156
VERSION     CD928156.1  GI:32775920
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 677)
  AUTHORS   Genoplante.
  TITLE     Genoplante, a major partnership french program in plant genomics
  JOURNAL   Unpublished (2003)
COMMENT,    Contact: Genoplante
            Genoplante
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplante' (http://www.genoplante.com
            and http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1..677
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="recital"
                     /db_xref="taxon:4565"
                     /clone="GR45104C10"
                     /tissue_type="grain (45 degrees per day after
                     pollination)"
                     /clone_lib="GR45"
ORIGIN

Alignment Scores:
Pred. No.:              3.13e-09      Length:        677
Score:                  168.00        Matches:       39
Percent Similarity:     49.71%        Conservative:  47
Best Local Similarity:  22.54%        Mismatches:    73
Query Match:            5.58%         Indels:        14
DB:                     14            Gaps:          2

US-09-913-878A-2_COPY_710_1282 (1-573) x CD928156 (1-677)

Qy        155 LeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPhe 174
               |||      :::      ::::::    :::|||    :::    ::::::        |||:::
Db         45 CTGGGGAAAATTGGTAGTGCAGAAACAAAGACAGCATCGAAAATTTTACTGCAAGGCTAT 104

Qy        175 AspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCys 194
               :::|||      :::    ||||||       |||                |||        :::
Db        105 GAACCAAGTTTAGAGCCTTACCTGTTGATGATTCTTAAGGCCCATCAGGATAATAGGCTG 164

Qy        195 AspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIle 214
                   ::::::::::::          |||    |||    :::           :::
Db        165 ACTGACATAAGAACTAGATGTAAGATTCATGTTCCAAAAGGCCGTGTTCTTATTGGTTGT 224

Qy        215 AlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPhe 234
                 |||       |||    ||||||       :::|||::::::       :::
Db        225 TTGGATGAAACAGGTGAATTAGAATATGGTCAAGTGTACATCAGAATTACAAAGAACAGC 284

Qy        235 ArgAspGluGluGluSerPhe------------------------------------Thr 242
               ::::::::::::::::::                                          |||
Db        285 AAAGAGCAGAAGGATAATTGTCAGCCATATTTGCTGAAGATAATGGAAAGAGAAAACA 344

Qy        243 LeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIle 262
                   :::            |||   |||::::::::|||           |||    ||||||
Db        345 GCAGTTGTTGTTGGAAAAGTTGCAGTATCGAAAAATCCTTGTCTCCATCCTGGTGACATC 404

Qy        263 GlnArgValArgAlaValPheLysProGluLeuHis------SerLeuLysAspValIle 280
```

```
               :::   :::   ||||||:::          |||:::        :::|||    |||    :::
Db        405 AGAGTACTTGAAGCTGTCTATGACCATGGTTTGTACGCTAAGAACCTGGTTGATTGTGTT 464

Qy        281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 300
               :::|||       :::|||:::    |||           :::    ||||||||||||   |||
Db        465 GTCTTCCCCCAAAGAGGAGAAAGGCCTCATCCAAATGAATGCTCCGGGGGCGATTTGGAC 524

Qy        301 GlyAspMetAlaTrpValCysTrpAspProGluIleVal 313
               ||||||:::    ::::::    ||||||    :::::::::
Db        525 GGTGACCTCTATTTTATCACTTGGGATGAGAAACTGATT 563

RESULT 14
CG223171
LOCUS       CG223171                 916 bp    DNA     linear   GSS 22-AUG-2003
DEFINITION  OGWAL14TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0519D04,
            genomic survey sequence.
ACCESSION   CG223171
VERSION     CG223171.1  GI:34123059
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 916)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OGWAL14TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..916
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0519D04"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

Alignment Scores:
Pred. No.:              5.98e-09      Length:        916
Score:                  167.50        Matches:       59
Percent Similarity:     39.85%        Conservative:  47
Best Local Similarity:  22.18%        Mismatches:    101
Query Match:            5.56%         Indels:        59
DB:                     29            Gaps:          8

US-09-913-878A-2_COPY_710_1282 (1-573) x CG223171 (1-916)

Qy         66 LysHisGlnArgThrLeuGluValArgSerValAlaSer------GluLeuLysSerAla 83
               |||    |||    |||        |||    |||              :::    :::    |||
Db         18 AAGTTCCAGTCCACGCACACCGTGCTGGAAGTGGTGTCCTGGACCAAGTTCCAGCCAGCG 77

Qy         84 GlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLysMet 103
                  ||||||   |||::::::      :::|||          :::
Db         78 TTCTTGAACCGCCAGATCATAACGCTGCTCACCACCCAAGGCGTTCCCGACGACGCCTTC 137

Qy        104 ArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLys 123
                  |||    :::    :::    ::::::        |||:::|||        |||:::
Db        138 TGGCAG---CTGCAGGAAGCGATGCTAGGCAAGCTCAGGCGGGTTTTGTCGGACAGCGAT 194
```

```
Qy       124 HisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArg 143
                 |||                                 ||||||       ::::::
Db       195 GTGGCG------------------------------TACGAGGTCGTGACCAACTCT 221

Qy       144 AlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGlu 163
                                                                     |||
Db       222 TGC------------------------------------------CCTGAGCAAGGA 236

Qy       164 GluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGln 183
               |||          ::::::   :::|||:::
Db       237 AGCACCGCAGGGCTGATGCTGGGCGCTGGCTAC--------------------------- 269

Qy       184 AspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIle 203
                         |||            |||:::|||             :::|||
Db       270 ---------TGGCCATCTGCACCTCGCAGATGCAGGGCCTTCTCAACAAAATGTGGATCT 320

Qy       204 ArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGlu 223
                      |||             :::    ::::::        |||    ||||||:::
Db       321 TTGTGCCCAAGGGAAGGTGGCTGATGGGCTGTCTCGACGAGTTTGGGATCCCTTGAGCAG 380

Qy       224 AsnGluVal--------------------HisValGly-------------------- 229
                :::                          |||||
Db       381 GGGCAGTGCTTTCTCCGGGTCTCGACGCCATCAGTCGGCAGCCGCTGCGTGAACCGTTCT 440

Qy       230 ------PheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCys 247
                     |||    |||:::::::::   :::    ::::::         :::
Db       441 CCGATATTTCCCTCGGAGTATAAGGGCGATGCCGATGCACAGGTGATTACTGGT------ 494

Qy       248 AspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAla 267
                |||::::::|||::::::|||            |||    |||::::::   :::   |||
Db       495 ACGGTTGTGATGGCGAAGAACCCATGCCTTCATCCAGGGGACGTCCGTATCCTTGAAGCT 554

Qy       268 ValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAsp 287
              |||      |||:::|||:::   :::    |||         :::|||         ||||||:::
Db       555 GTGGATGTCCCTGATCTGTATCACATTGTCGACTGCTTCGTCTTCCCCAAGAAAGGTGAG 614

Qy       288 ValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCys 307
                |||    |||    |||    ||||||    |||    |||||||||:::    :::|||
Db       615 AGGCCACACGCCGACAAAGCTTCCGGGAGCGACCTCGACGGCGACGTCTACTTCGTGACG 674

Qy       308 TrpAspProGluIleVal 313
              ||||||      :::|||
Db       675 TGGGACGAGAGCCTTGTA 692
```

RESULT 15
CF134248

```
LOCUS       CF134248                722 bp    mRNA    linear   EST 24-JUL-2003
DEFINITION  WHE4369_F09_L17ZT Wheat meiotic floret cDNA library Triticum
            aestivum cDNA clone WHE4369_F09_L17, mRNA sequence.
ACCESSION   CF134248
VERSION     CF134248.1  GI:33219166
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 722)
  AUTHORS   Anderson,O.D., Chao,S., Crossman,C., Langridge,P., Lazo,G.R.,
            Pham,J., Rausch,C.J., Sutton,T., Woo,J. and Wilson,C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Meiotic floret cDNA library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1..722
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE4369_F09_L17"
                     /tissue_type="Whole florets with anthers"
                     /dev_stage="Meiotic stages pre-meiosis-metaphase I"
                     /lab_host="E. coli DH10B"
                     /clone_lib="Wheat meiotic floret cDNA library"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Plants
                     were grown in a glasshouse. Anther meiotic stage was
                     determined microscopically after removing a single anther
                     from a primary floret. If determined to be between (and
                     including) meiotic stages pre-meiosis and metaphase I,
                     remaining floret tissue was collected and pooled for
                     library construction. The tissue, total RNA, and poly(A)
                     RNA were prepared, cDNA synthesised, and directionally
                     ligated into pSPORT1 by Tim Sutton in the P Langridge Lab
                     at the Department of Plant Science, University of
                     Adelaide, Waite Campus, Australia. Average insert size
                     1.4Kb. Plasmid DNA preparations and DNA sequencing were
                     performed in the OD Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:              6.12e-09        Length:         722
Score:                  166.00          Matches:        51
Percent Similarity:     45.54%          Conservative:   46
Best Local Similarity:  23.94%          Mismatches:     102
Query Match:            5.51%           Indels:         14
DB:                     14              Gaps:           4

US-09-913-878A-2_COPY_710_1282 (1-573) x CF134248 (1-722)

Qy       115 LeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgPro---ValGluPheArg 133
              |||   ||||||                    :::||||||    |||          |||
Db        42 TTAAACCGTCAGATCATAACATTACTCTCCTCCTTGAATGTCCCTGATGCTATCTTTTCA 101

Qy       134 GlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValPro 153
              |||        :::      |||:::               ::::::         |||
Db       102 CAAATGCAGAAAGACATGCTGTCTAATCTCAACAATATTTGACTGACACTGATGTTGCT 161

Qy       154 PheLeuAlaGlyLeuProAspSerGlnGluGlu------ThrLeuAsnPheLeuMetAsn 171
              |||                  :::|||      |||
Db       162 TTTGACGTTGTAACCACCTCTTGTGCTGACGAAGGAAACACCGCAGCACTGATGTTGAGT 221

Qy       172 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 191
             :::|||       |||      :::   :::|||:::              ::::::
Db       222 GCTGGCATTTCACCTGGAACTGAGCCGCACCTGAAAGCATTGCTGTTGGCTATAAGGTCC 281

Qy       192 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 211
                :::      |||      |||      |||   |||     :::      ::::::
Db       282 TCACAGCTACTGGGTCTTTTGGAAAAATCAAGGATTTTGTTCCCAAGGGAAGGTGGTTG 341

Qy       212 TyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer 231
                      |||      |||:::||||||:::     :::          :::     |||
Db       342 ATGGGCTGCCTTGATGAACTTGGGATCCTTGAGCAAGGACAGTGCTTTATTCGGGCCTCA 401

Qy       232 SerLys-----------------PheArgAspGluGluGluSerPheThrLeuLeuSer 245
             |||                   :::          |||   :::        :::
Db       402 TCTCCAGTACTCAATAATTCTCTGCTGAAGCATGGACCAAGATCTTCCTCAGCGAACAAC 461

Qy       246 AspCysAsp---------------ValLeuValAlaArgSerProAlaHisPheProSer 260
             :::   :::                  |||::::::|||::::::|||         |||
Db       462 AATGCAGAGACCGTTAAAGGTACTGTTGTTATGGCAAAGAATCCGTGCCTTCATCCAGGG 521

Qy       261 AspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIle 280
             |||::::::   :::   |||:::         |||   ||||||   |||   |||   :::
Db       522 GATGTCCGGATCCTTGAAGCCATTGATGTGCCTGCTTTGCATCACCTTGTTGATTGCTTG 581

Qy       281 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 300
```

```
                 :::|||          |||:::   |||   |||   :::   ||||||   |||   |||
Db          582 GTCTTTCCCAAGAATGGCGAGAGGCCACACGCCAATGAAGCGTCTGGGAGTGATCTTGAT 641

Qy          301 GlyAspMetAlaTrpValCysTrpAspProGluIleVal 313
                ||||||:::   :::|||   ||||||   :::::::::
Db          642 GGGGATCTATACTTTGTGACGTGGGATGAAAAACTTATC 680

Search completed: March 31, 2004, 13:00:01
Job time : 2240.7 secs
```

```
OM nucleic - nucleic search, using sw model

Run on:         March 30, 2004, 15:33:39 ; Search time 19844 Seconds
                                           (without alignments)
                                           17571.795 Million cell updates/sec

Title:          US-09-913-878A-1
Perfect score:  8045
Sequence:       1 ggtaccgggccccccctcga..........cgtattacaatcccagtcac 8045

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:      6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	8045	100.0	8045	6	AX034425	AX034425 Sequence
	2	4206	52.3	4206	8	NCR133528	AJ133528 Neurospor
	3	284.2	3.5	333321	3	AC116986	AC116986 Dictyoste
	4	276.6	3.4	252215	2	AC095632	AC095632 Rattus no
	5	273.4	3.4	229259	2	AC096212	AC096212 Rattus no
	6	273.2	3.4	200412	2	AC115666	AC115666 Rattus no
c	7	273.2	3.4	222913	2	AC120325	AC120325 Rattus no
	8	271.4	3.4	254733	3	AC117075	AC117075 Dictyoste
c	9	271	3.4	118689	2	AC141123	AC141123 Rattus no
c	10	271	3.4	205454	10	AC096792	AC096792 Rattus no
	11	269.2	3.3	198573	2	AC110118	AC110118 Rattus no
	12	268.6	3.3	255678	2	AC099432	AC099432 Rattus no
	13	268.4	3.3	279242	2	AC114079	AC114079 Rattus no
	14	268.4	3.3	346208	2	AC128290	AC128290 Rattus no
c	15	267.8	3.3	211295	2	AC130618	AC130618 Rattus no
c	16	267.2	3.3	110000	3	AC116984_0	AC116984 Dictyoste
c	17	267.2	3.3	110000	3	AC116984_1	Continuation (2 of
	18	266.6	3.3	240214	2	AC126871	AC126871 Rattus no
	19	266.4	3.3	175351	2	AC117900	AC117900 Rattus no
	20	266.4	3.3	247341	2	AC099178	AC099178 Rattus no
	21	264.2	3.3	136240	3	AC117070	AC117070 Dictyoste
	22	263.4	3.3	7921	3	AF153362	AF153362 Dictyoste
	23	261.6	3.3	110000	2	AC129796_0	AC129796 Rattus no
c	24	261.6	3.3	240800	2	AC095964	AC095964 Rattus no
c	25	261.6	3.3	274697	2	AC098240	AC098240 Rattus no
	26	261.2	3.2	282355	2	AC104401	AC104401 Rattus no
	27	260.6	3.2	256528	2	AC108270	AC108270 Rattus no
	28	260.4	3.2	177054	2	AC126897	AC126897 Rattus no
c	29	260.4	3.2	232357	2	AC127133	AC127133 Rattus no
c	30	260.2	3.2	196662	2	AC129737	AC129737 Rattus no
	31	260.2	3.2	219107	2	AC115259	AC115259 Rattus no
c	32	260	3.2	96573	2	AC136672	AC136672 Rattus no
c	33	260	3.2	237900	2	AC132525	AC132525 Rattus no
c	34	260	3.2	245524	2	AC102954	AC102954 Rattus no
	35	259.2	3.2	301081	2	AC108658	AC108658 Rattus no
	36	259.2	3.2	347444	2	AC111888	AC111888 Rattus no
	37	259	3.2	253094	2	AC113774	AC113774 Rattus no
	38	258.2	3.2	217746	2	AC112302	AC112302 Rattus no
	39	257.8	3.2	104708	2	AC139336	AC139336 Rattus no
	40	257.8	3.2	241531	2	AC122959	AC122959 Rattus no
	41	257.8	3.2	241765	2	AC132557	AC132557 Rattus no
c	42	257.8	3.2	256879	3	AC116982	AC116982 Dictyoste
	43	257.8	3.2	278876	2	AC120936	AC120936 Rattus no
	44	257	3.2	232373	2	AC134805	AC134805 Rattus no
c	45	256.8	3.2	264726	2	AC096844	AC096844 Rattus no

ALIGNMENTS

```
RESULT 1
AX034425
LOCUS       AX034425                 8045 bp    DNA     linear   PAT 22-SEP-2000
DEFINITION  Sequence 1 from Patent WO0050581.
ACCESSION   AX034425
VERSION     AX034425.1  GI:10303130
KEYWORDS    .
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Cogoni,C. and Macino,G.
  TITLE     Isolation and characterization of a n. Crassa silencing gene and
            uses thereof
```

```
JOURNAL     Patent: WO 0050581-A 1 31-AUG-2000;
            UNIV ROMA (IT) ; COGONI CARLO (IT) ; MACINO GIUSEPPE (IT)
FEATURES             Location/Qualifiers
     source          1..8045
                     /organism="Neurospora crassa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5141"
     CDS             2447..6655
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC10121.1"
                     /db_xref="GI:10303131"
                     /db_xref="REMTREMBL:CAC10121"
                     /translation="MNPITPRKRNSPVEEIINRLNNDYNLGLQCVADTTLTPHRRKEL
                     AESDEDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLP
                     WSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNT
                     KRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQC
                     TKRAKGKLSDNVAAAAAPPVPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQ
                     GTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKS
                     HIDNQPLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAP
                     VVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDV
                     TDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTA
                     PLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQ
                     WLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFRP
                     DVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKTYAI
                     MTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFG
                     SAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQL
                     LPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRV
                     SHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRV
                     GRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQ
                     RVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEM
                     PLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNY
                     KERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDP
                     MYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWDPDLA
                     SYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQVY
                     EKWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHKSPKF
                     VWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGDGSEYP
                     DPEVYEVLGDDDFDGIGFTGNGDY"
ORIGIN

  Query Match          100.0%;  Score 8045;  DB 6;  Length 8045;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 8045;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCTGAAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCTGAAAA 60

Qy         61 GTCCTAAAGTCTCCTGTCTCTCATGCTAGGTAGCCCTACTGCGTGTGTAGTGTAGTAGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTCCTAAAGTCTCCTGTCTCTCATGCTAGGTAGCCCTACTGCGTGTGTAGTGTAGTAGGT 120

Qy        121 CCGAATAGGTTCGTACGCTGTCCCCTTTAGGGGGCTGGTTTAACAGGGAGACAGGGTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGAATAGGTTCGTACGCTGTCCCCTTTAGGGGGCTGGTTTAACAGGGAGACAGGGTACA 180

Qy        181 GTCTCTCGGGAACAGAGACGACCAAGCTTTGTTGTGTGACATCATCTTGAGACCATTCGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCTCTCGGGAACAGAGACGACCAAGCTTTGTTGTGTGACATCATCTTGAGACCATTCGT 240

Qy        241 CTGAGGGTCCATACAACCATACATGCTGTGTAGAGAGTGGTGGGAATATTAGAGGTTACT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGAGGGTCCATACAACCATACATGCTGTGTAGAGAGTGGTGGGAATATTAGAGGTTACT 300

Qy        301 TTGAGAGTTTCTGAAGAGATTGAGACTGTAGGTACGTTCATTTATCGGTGCGACGAGAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGAGAGTTTCTGAAGAGATTGAGACTGTAGGTACGTTCATTTATCGGTGCGACGAGAAT 360

Qy        361 GGGCACGCATGAGTCGACATTTGCAATTTGTTGGTTGGCACGCACATCGATAGTGTACAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGGCACGCATGAGTCGACATTTGCAATTTGTTGGTTGGCACGCACATCGATAGTGTACAC 420

Qy        421 ATGTGTCGGCCTGAGACTGTCCAATGGACCAATGGTAAGGCGTGAGCGGCGAGTGGGGTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATGTGTCGGCCTGAGACTGTCCAATGGACCAATGGTAAGGCGTGAGCGGCGAGTGGGGTC 480

Qy        481 AACTTGAGGTTAATGTACACAGTACTTGAGGCTGTAGCGGGCAGCGGAAGCATCCGTCCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AACTTGAGGTTAATGTACACAGTACTTGAGGCTGTAGCGGGCAGCGGAAGCATCCGTCCT 540

Qy        541 CGAATGGACATTCCCAAGCCAGCCAACACTGGTGTCCATTTAAAAGCGTTTTCTCCATGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGAATGGACATTCCCAAGCCAGCCAACACTGGTGTCCATTTAAAAGCGTTTTCTCCATGG 600

Qy        601 GGGAACAATGGCCGGCGTAGTATCTGTAAGAGGACCTAGAGGACGGTATTAGAACATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGGAACAATGGCCGGCGTAGTATCTGTAAGAGGACCTAGAGGACGGTATTAGAACATTTG 660

Qy        661 GGTGATGTATCTAGATCACCGATGTGTCCCTGAATAGTCCATGAATAGTCTGGGATGAGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGTGATGTATCTAGATCACCGATGTGTCCCTGAATAGTCCATGAATAGTCTGGGATGAGT 720

Qy        721 CCGAGATCAGTCCGGGATGTGTCACCGATATATCCATGACTTATCTCTGATTTATCCCAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCGAGATCAGTCCGGGATGTGTCACCGATATATCCATGACTTATCTCTGATTTATCCCAG 780

Qy        781 CGAGACCCCAATCAGTATGACCGACAGCGACACCATCACCGGCGCCAGCGTATGAGATCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGAGACCCCAATCAGTATGACCGACAGCGACACCATCACCGGCGCCAGCGTATGAGATCC 840

Qy        841 TGAACATATGTTAGGCATATGTATGGCATATGTATGCATAGAGCATGCACCACATATCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGAACATATGTTAGGCATATGTATGGCATATGTATGCATAGAGCATGCACCACATATCTG 900

Qy        901 AGAGCTGCCTGTGACTACCTAGAGGTATGTAGAGGTAATGTCACGTACCTTACCTCTACC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTGCCTGTGACTACCTAGAGGTATGTAGAGGTAATGTCACGTACCTTACCTCTACC 960

Qy        961 AAATGTATCCCGCAGTCACTCGTATTCCCGCTGTAGTGTACACTAGAACTAGCCCCGCCT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAATGTATCCCGCAGTCACTCGTATTCCCGCTGTAGTGTACACTAGAACTAGCCCCGCCT 1020

Qy       1021 TTTGGATCCTGGAGTGTCCCCATTCCACGCACTTGCCAAGAACCACAACCTCGCAAAGCG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTGGATCCTGGAGTGTCCCCATTCCACGCACTTGCCAAGAACCACAACCTCGCAAAGCG 1080

Qy       1081 CGCTGACAGGGCACCGTGCTGCCTGGACCTTTGCGGGCCGTTCCTTCTTGACCGGCGCCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCTGACAGGGCACCGTGCTGCCTGGACCTTTGCGGGCCGTTCCTTCTTGACCGGCGCCA 1140

Qy       1141 GGCAGCAGACCGGTTGGTTCTTGTTCCTCAAGGTGACGAAGGCGATTAAAAAAATCCAAGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GGCAGCAGACCGGTTGGTTCTTGTTCCTCAAGGTGACGAAGGCGATTAAAAAAATCCAAGA 1200

Qy       1201 CAACCAGATCCTGGAACCTTTGATCATCCCTCGTTTACTCAGACGAAAAAACTAAAAAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CAACCAGATCCTGGAACCTTTGATCATCCCTCGTTTACTCAGACGAAAAAACTAAAAAAC 1260

Qy       1261 TTCAAAGTCGTTGTCAGATTAATTGAACCGTTAAAACAGGACATGTCCGCCCTTTCGTCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCAAAGTCGTTGTCAGATTAATTGAACCGTTAAAACAGGACATGTCCGCCCTTTCGTCT 1320

Qy       1321 CTCATGCCCTCGAAACCTCACGTCTTCCATCTCTCGTCCGATCCCCCCCTCCCCACGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCATGCCCTCGAAACCTCACGTCTTCCATCTCTCGTCCGATCCCCCCCTCCCCACGACC 1380

Qy       1381 CTGATTCAGCAGCGAACTGACCTGTCACTTTACTGTGGACTCCTCCTAGAGTAGGGCTGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGATTCAGCAGCGAACTGACCTGTCACTTTACTGTGGACTCCTCCTAGAGTAGGGCTGG 1440

Qy       1441 CCATCCCCGGGCCCCCTGCAGCTTCTGCAAGTCAAAGCGATGGAAGGATGGCCGTCCAGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCATCCCCGGGCCCCCTGCAGCTTCTGCAAGTCAAAGCGATGGAAGGATGGCCGTCCAGT 1500

Qy       1501 GGGTGCGGCATACAGCCATTTCCAAAATGGTCTCGTCGCATGCGGAAGAAAAAAAATGGA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1501 GGGTGCGGCATACAGCCATTTCCAAAATGGTCTCGTCGCATGCGGAAGAAAAAAAATGGA 1560

Qy   1561 AGGCTACGAGCCTGCGACACCTGTACGGTCACTGCGGCTACCCTTCACCACCACCAGTGA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AGGCTACGAGCCTGCGACACCTGTACGGTCACTGCGGCTACCCTTCACCACCACCAGTGA 1620

Qy   1621 TATCATCACCCATACCATCAGGGCCACTACCGCCGCGCTCTCTCGAACATCGGGCTGAAT 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TATCATCACCCATACCATCAGGGCCACTACCGCCGCGCTCTCTCGAACATCGGGCTGAAT 1680

Qy   1681 GTGACCAATTGCGCCGACAGCGGGCTTTTCGATCCTGTGGGCAGCTTCTTCCACCGGTTT 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GTGACCAATTGCGCCGACAGCGGGCTTTTCGATCCTGTGGGCAGCTTCTTCCACCGGTTT 1740

Qy   1741 GCGCTGATCTGGGAGCGTATCGAGCTGGATTGGGACTGTTTACCACTCGAATCACTTGAC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GCGCTGATCTGGGAGCGTATCGAGCTGGATTGGGACTGTTTACCACTCGAATCACTTGAC 1800

Qy   1801 GGCTCGGCGCGATTCTATGGCCATCCACCGCGATTGATCGGTCCCTTCACTGTTTCCTGC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 GGCTCGGCGCGATTCTATGGCCATCCACCGCGATTGATCGGTCCCTTCACTGTTTCCTGC 1860

Qy   1861 AACCTACTCTCCCGACCAGATCAGACGAGGCTTTATGCCCAGTTCCGGGCGTTGCCAATT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AACCTACTCTCCCGACCAGATCAGACGAGGCTTTATGCCCAGTTCCGGGCGTTGCCAATT 1920

Qy   1921 ATAGCCACAGCACGAAGCTCGCCAAGTTCGTTCGAAATGCTCAAAACTGTTCACACACTT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ATAGCCACAGCACGAAGCTCGCCAAGTTCGTTCGAAATGCTCAAAACTGTTCACACACTT 1980

Qy   1981 TGCACGAACAGGACCACCACGAAAAACCACCTTGGCCCGCTCGCGCGAACCCGTGTCTTT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 TGCACGAACAGGACCACCACGAAAAACCACCTTGGCCCGCTCGCGCGAACCCGTGTCTTT 2040

Qy   2041 GAACTCGATCAGCCCCAATCCCAAACCAGGTCGCCTTTCCCCGCTGCGCCTGCTTGTCCC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 GAACTCGATCAGCCCCAATCCCAAACCAGGTCGCCTTTCCCCGCTGCGCCTGCTTGTCCC 2100

Qy   2101 GACTTCGTGAAGCAAACAAGACTGGCGGCTATGGCCAAGGCCCATTCCATGACAATTGGT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GACTTCGTGAAGCAAACAAGACTGGCGGCTATGGCCAAGGCCCATTCCATGACAATTGGT 2160

Qy   2161 TAATCGAGGTCAAAAACTACAAGAGATCAACGCCGCAAAGATATACTTGTCGTTCACCCC 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TAATCGAGGTCAAAAACTACAAGAGATCAACGCCGCAAAGATATACTTGTCGTTCACCCC 2220

Qy   2221 GACAACCGCAAAGCCTTGAAGCGCAAAGCCACAGTGGTGTTGTTTACTTCCGCACACTTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GACAACCGCAAAGCCTTGAAGCGCAAAGCCACAGTGGTGTTGTTTACTTCCGCACACTTA 2280

Qy   2281 TCGACCCTTACGCCCAAGACGCTCCCGCTTGGTAGAAGCTCGAGTCGGGTAATCAGGTAT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCGACCCTTACGCCCAAGACGCTCCCGCTTGGTAGAAGCTCGAGTCGGGTAATCAGGTAT 2340

Qy   2341 ATCTCACAGTCGCCTTGTCCTCCTTTTCGGTGGTGGTCGCTTTTATCATGCCTTATCGCT 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 ATCTCACAGTCGCCTTGTCCTCCTTTTCGGTGGTGGTCGCTTTTATCATGCCTTATCGCT 2400

Qy   2401 AACCCGGTTGTATACTCTACTCCAGGCCCGATCTCCCACGGCACAAATGAACCCTATTAC 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 AACCCGGTTGTATACTCTACTCCAGGCCCGATCTCCCACGGCACAAATGAACCCTATTAC 2460

Qy   2461 TCCTAGGAAGAGGAATAGCCCCGTCGAGGAAATCATAAACCGGCTCAATAACGACTACAA 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TCCTAGGAAGAGGAATAGCCCCGTCGAGGAAATCATAAACCGGCTCAATAACGACTACAA 2520

Qy   2521 CCTGGGCCTCCAGTGTGTCGCAGACACAACTCTCACCCCCCACCGCCGGAAGGAGCTGGC 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CCTGGGCCTCCAGTGTGTCGCAGACACAACTCTCACCCCCCACCGCCGGAAGGAGCTGGC 2580

Qy   2581 CGAGAGTGACGAGGATTTCGGTCGCCATGACAAGATCTACAGAGCCCTGAACTTTCTCTA 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CGAGAGTGACGAGGATTTCGGTCGCCATGACAAGATCTACAGAGCCCTGAACTTTCTCTA 2640

Qy   2641 CTGGCGGAAGGATGACTCCCTGAACCAGGCAGAAGCCAACTTCTTCATCGAGGCCAAAGC 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 CTGGCGGAAGGATGACTCCCTGAACCAGGCAGAAGCCAACTTCTTCATCGAGGCCAAAGC 2700

Qy   2701 TGCGAGCTCGAACTGGGTGCCCAAAGCCCACGCCGACCCTGACACGCTTCCGTGGTCCAA 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 TGCGAGCTCGAACTGGGTGCCCAAAGCCCACGCCGACCCTGACACGCTTCCGTGGTCCAA 2760

Qy   2761 GGAACCTCCCCGCGCCGCTACTGCCGGCCAACAATGGGCATTGCAGACTGTGTTGCTCGA 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 GGAACCTCCCCGCGCCGCTACTGCCGGCCAACAATGGGCATTGCAGACTGTGTTGCTCGA 2820

Qy   2821 GGTGCTTAATAGGTTTATGCCACCTCCCAATAACACACCAGGTCGAACGTTTGGCAGAAC 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GGTGCTTAATAGGTTTATGCCACCTCCCAATAACACACCAGGTCGAACGTTTGGCAGAAC 2880

Qy   2881 TCTAAGCGGCCCAAGTGGCCTGAGCCGCCCAACCTCTACCAACACCAAACGCAAGGATGA 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TCTAAGCGGCCCAAGTGGCCTGAGCCGCCCAACCTCTACCAACACCAAACGCAAGGATGA 2940

Qy   2941 GCCCGCCAATGTCACTTTCGCTGATCCGCCCAAACGCTCGTTGACTCGCTCTGCCACAGG 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 GCCCGCCAATGTCACTTTCGCTGATCCGCCCAAACGCTCGTTGACTCGCTCTGCCACAGG 3000

Qy   3001 TCCTCCTATTCACGGCGCGGCGATACCCCTAAAGTTCCCCGATCCAGTGAATACCGGTTC 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 TCCTCCTATTCACGGCGCGGCGATACCCCTAAAGTTCCCCGATCCAGTGAATACCGGTTC 3060

Qy   3061 CAAACGACCATCTCTCGAGAGTGAGAATCTCAATCAGTGCACCAAGCGGGCCAAGGGCAA 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 CAAACGACCATCTCTCGAGAGTGAGAATCTCAATCAGTGCACCAAGCGGGCCAAGGGCAA 3120

Qy   3121 GCTGTCTGATAATGTTGCCGCTGCCGCCGCCCCGCCCGTGCCTATTGCGAGCGCTTTGGA 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 GCTGTCTGATAATGTTGCCGCTGCCGCCGCCCCGCCCGTGCCTATTGCGAGCGCTTTGGA 3180

Qy   3181 CAAGGTACCGACTCGAAGGCATGCCAATACGAGAGATCCCACGGCGACAGGTCATAGACG 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 CAAGGTACCGACTCGAAGGCATGCCAATACGAGAGATCCCACGGCGACAGGTCATAGACG 3240

Qy   3241 AGCGGACCAGGTGGATTCCTTTGATACATCTCAAGGCACTTCCTATGGTTCGAGTGTCTT 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 AGCGGACCAGGTGGATTCCTTTGATACATCTCAAGGCACTTCCTATGGTTCGAGTGTCTT 3300

Qy   3301 CAGCGCTTGCCGTCACAATCAGAGCACTACCCAGAGTAGTTTTGAGGCTCCTCCTTCACA 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CAGCGCTTGCCGTCACAATCAGAGCACTACCCAGAGTAGTTTTGAGGCTCCTCCTTCACA 3360

Qy   3361 GCCCAGAGAGAAGCGGCCTGTGGATGCCACGGTCTTTGAGGCTGGACACTTGATTGAGTC 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 GCCCAGAGAGAAGCGGCCTGTGGATGCCACGGTCTTTGAGGCTGGACACTTGATTGAGTC 3420

Qy   3421 TCCTAGCAAAGGAAGAACAACCAAGTCCCACATAGATAACCAGCCCCTTTCATCGTCTTC 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 TCCTAGCAAAGGAAGAACAACCAAGTCCCACATAGATAACCAGCCCCTTTCATCGTCTTC 3480

Qy   3481 CCAGGGTGAAACTTCGTTCAGCACTTACTATGAGTCGTTTCCAAGTTCCGGCGGCGAGGG 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CCAGGGTGAAACTTCGTTCAGCACTTACTATGAGTCGTTTCCAAGTTCCGGCGGCGAGGG 3540

Qy   3541 CGCAATTCCCGAGCCGAGTCGCTCAAATGGACTGGCTCGGAGCGAAGAAAGCGCTCGATC 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 CGCAATTCCCGAGCCGAGTCGCTCAAATGGACTGGCTCGGAGCGAAGAAAGCGCTCGATC 3600

Qy   3601 TCAGGTTCAAGTTCATGCCCCGGTGGTTGCAGCTCGGCTGAGAAATATTTGGCCGAAATT 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 TCAGGTTCAAGTTCATGCCCCGGTGGTTGCAGCTCGGCTGAGAAATATTTGGCCGAAATT 3660

Qy   3661 TCCCAAATGGCTACACGAAGCTCCTCTCGCTGTTGCATGGGAAGTTACCAGACTCTTTAT 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 TCCCAAATGGCTACACGAAGCTCCTCTCGCTGTTGCATGGGAAGTTACCAGACTCTTTAT 3720
```

```
Qy       3721 GCACTGCAAAGTAGACTTGGAAGACGAGAGCCTGGGCCTAAAGTACGACCCTTCCTGGTC 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 GCACTGCAAAGTAGACTTGGAAGACGAGAGCCTGGGCCTAAAGTACGACCCTTCCTGGTC 3780

Qy       3781 TACCGCGCGCGATGTCACAGATATCTGGAAGACTCTCTACCGGCTTGATGCTTTCCGTGG 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 TACCGCGCGCGATGTCACAGATATCTGGAAGACTCTCTACCGGCTTGATGCTTTCCGTGG 3840

Qy       3841 TAAACCCTTTCCAGAAAAGCCGCCCAACGACGTGTTCGTGACCGCAATGACGGGCAACTT 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 TAAACCCTTTCCAGAAAAGCCGCCCAACGACGTGTTCGTGACCGCAATGACGGGCAACTT 3900

Qy       3901 TGAGAGCAAAGGTAGTGCCGTTGTTCTCTCTGCTGTTCTAGACTACAATCCGGACAACTC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 TGAGAGCAAAGGTAGTGCCGTTGTTCTCTCTGCTGTTCTAGACTACAATCCGGACAACTC 3960

Qy       3961 GCCTACTGCGCCCCTTTACCTTGTGAAGCTGAAGCCGCTCATGTTCGAGCAGGGCTGTCG 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 GCCTACTGCGCCCCTTTACCTTGTGAAGCTGAAGCCGCTCATGTTCGAGCAGGGCTGTCG 4020

Qy       4021 ACTCACCCGTCGGTTCGGTCCTGATAGGTTTTTCGAGATCCTTATACCGTCGCCTACGAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 ACTCACCCGTCGGTTCGGTCCTGATAGGTTTTTCGAGATCCTTATACCGTCGCCTACGAG 4080

Qy       4081 CACCAGCCCAAGTGTACCGCCGGTGGTCAGCAAACAACCAGGTGCGGTCGAAGAAGTCAT 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 CACCAGCCCAAGTGTACCGCCGGTGGTCAGCAAACAACCAGGTGCGGTCGAAGAAGTCAT 4140

Qy       4141 CCAGTGGCTCACGATGGGGCAACATTCTCTTGTAGGCCGCCAATGGCGCGCTTTCTTCGC 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 CCAGTGGCTCACGATGGGGCAACATTCTCTTGTAGGCCGCCAATGGCGCGCTTTCTTCGC 4200

Qy       4201 CAAAGATGCCGGATACAGGAAACCTCTCAGGGAGTTCCAGCTCCGCGCCGAGGACCCGAA 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 CAAAGATGCCGGATACAGGAAACCTCTCAGGGAGTTCCAGCTCCGCGCCGAGGACCCGAA 4260

Qy       4261 ACCCATCATCAAGGAGAGAGTCCACTTCTTTGCCGAGACCGGCATTACGTTCCGACCTGA 4320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 ACCCATCATCAAGGAGAGAGTCCACTTCTTTGCCGAGACCGGCATTACGTTCCGACCTGA 4320

Qy       4321 TGTGTTCAAGACGAGATCTGTCGTTCCGGCAGAGGAACCTGTAGAGCAACGGACCGAGTT 4380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4321 TGTGTTCAAGACGAGATCTGTCGTTCCGGCAGAGGAACCTGTAGAGCAACGGACCGAGTT 4380

Qy       4381 CAAAGTTAGTCAAATGCTGGACTGGCTCCTGCAACTCGACAACAACACTTGGCAGCCGCA 4440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4381 CAAAGTTAGTCAAATGCTGGACTGGCTCCTGCAACTCGACAACAACACTTGGCAGCCGCA 4440

Qy       4441 CCTCAAGTTGTTCTCCCGTATCCAGCTCGGTCTGAGTAAGACATATGCCATTATGACATT 4500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4441 CCTCAAGTTGTTCTCCCGTATCCAGCTCGGTCTGAGTAAGACATATGCCATTATGACATT 4500

Qy       4501 GGAGCCTCACCAGATCAGACACCACAAGACCGATCTTCTTTCACCTTCAGGCACTGGCGA 4560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4501 GGAGCCTCACCAGATCAGACACCACAAGACCGATCTTCTTTCACCTTCAGGCACTGGCGA 4560

Qy       4561 AGTGATGAATGACGGTGTAGGCCGCATGTCGCGAAGCGTGGCCAAGAGGATACGCGATGT 4620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 AGTGATGAATGACGGTGTAGGCCGCATGTCGCGAAGCGTGGCCAAGAGGATACGCGATGT 4620

Qy       4621 TCTCGGTTTGGGTGATGTGCCCTCTGCTGTGCAAGGGCGGTTTGGTTCGGCCAAGGGAAT 4680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4621 TCTCGGTTTGGGTGATGTGCCCTCTGCTGTGCAAGGGCGGTTTGGTTCGGCCAAGGGAAT 4680

Qy       4681 GTGGGTTATCGACGTTGACGACACAGGCGATGAGGATTGGATCGAGACATACCCGTCCCA 4740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4681 GTGGGTTATCGACGTTGACGACACAGGCGATGAGGATTGGATCGAGACATACCCGTCCCA 4740

Qy       4741 GCGCAAGTGGGAATGCGACTTCGTTGATAAACATCAACGTACCCTCGAAGTCCGGAGCGT 4800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4741 GCGCAAGTGGGAATGCGACTTCGTTGATAAACATCAACGTACCCTCGAAGTCCGGAGCGT 4800

Qy       4801 GGCTTCTGAACTGAAGTCAGCTGGTCTCAACCTACAGCTGTTACCTGTCCTGGAAGATAG 4860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4801 GGCTTCTGAACTGAAGTCAGCTGGTCTCAACCTACAGCTGTTACCTGTCCTGGAAGATAG 4860

Qy       4861 AGCCAGGGACAAGGTGAAGATGCGCCAGGCAATCGGTGACCGTCTTATCAACGATTTGCA 4920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4861 AGCCAGGGACAAGGTGAAGATGCGCCAGGCAATCGGTGACCGTCTTATCAACGATTTGCA 4920

Qy       4921 ACGACAGTTCAGCGAGCAAAAGCATGCTTTGAATCGCCCAGTGGAATTTCGCCAATGGGT 4980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4921 ACGACAGTTCAGCGAGCAAAAGCATGCTTTGAATCGCCCAGTGGAATTTCGCCAATGGGT 4980

Qy       4981 TTACGAGAGTTATTCCAGTCGCGCAACTCGAGTCAGCCACGGCCGTGTGCCTTTTCTTGC 5040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4981 TTACGAGAGTTATTCCAGTCGCGCAACTCGAGTCAGCCACGGCCGTGTGCCTTTTCTTGC 5040

Qy       5041 TGGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAACAGTGGGTTCGATCC 5100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5041 TGGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAACAGTGGGTTCGATCC 5100

Qy       5101 CAAGAAGCAAAAGTACTTGCAAGACATCGCCTGGGATCTTCAAAAGCGGAAATGTGACAC 5160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5101 CAAGAAGCAAAAGTACTTGCAAGACATCGCCTGGGATCTTCAAAAGCGGAAATGTGACAC 5160

Qy       5161 GTTGAAGTCCAAGCTGAACATCCGTGTCGGTCGATCAGCATACATTTACATGATTGCCGA 5220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5161 GTTGAAGTCCAAGCTGAACATCCGTGTCGGTCGATCAGCATACATTTACATGATTGCCGA 5220

Qy       5221 TTTCTGGGGTGTGCTTGAGGAAAATGAGGTTCATGTCGGATTCTCCTCAAAGTTCAGAGA 5280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5221 TTTCTGGGGTGTGCTTGAGGAAAATGAGGTTCATGTCGGATTCTCCTCAAAGTTCAGAGA 5280

Qy       5281 CGAGGAGGAGTCTTTTACACTCCTATCGGACTGTGATGTCCTCGTGGCGCGATCCCCAGC 5340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5281 CGAGGAGGAGTCTTTTACACTCCTATCGGACTGTGATGTCCTCGTGGCGCGATCCCCAGC 5340

Qy       5341 CCATTTCCCTAGTGATATCCAACGGGTTCGAGCAGTCTTCAAGCCAGAGCTCCACAGTCT 5400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5341 CCATTTCCCTAGTGATATCCAACGGGTTCGAGCAGTCTTCAAGCCAGAGCTCCACAGTCT 5400

Qy       5401 CAAGGATGTAATCATCTTCTCTACTAAAGGAGATGTACCGCTTGCTAAGAAGCTATCTGG 5460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5401 CAAGGATGTAATCATCTTCTCTACTAAAGGAGATGTACCGCTTGCTAAGAAGCTATCTGG 5460

Qy       5461 TGGAGACTACGACGGCGATATGGCCTGGGTCTGCTGGGATCCGGAGATCGTCGATGGTTT 5520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5461 TGGAGACTACGACGGCGATATGGCCTGGGTCTGCTGGGATCCGGAGATCGTCGATGGTTT 5520

Qy       5521 CGTCAATGCGGAAATGCCTCTGGAGCCCGACCTGTCTAGGTACCTAAAGAAGGACAAAAC 5580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5521 CGTCAATGCGGAAATGCCTCTGGAGCCCGACCTGTCTAGGTACCTAAAGAAGGACAAAAC 5580

Qy       5581 GACTTTCAAACAACTTATGGCCTCACACGGCACGGGCTCAGCGGCCAAAGAGCAGACTAC 5640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5581 GACTTTCAAACAACTTATGGCCTCACACGGCACGGGCTCAGCGGCCAAAGAGCAGACTAC 5640

Qy       5641 ATACGATATGATCCAGAAGAGCTTCCATTTCGCCCTGCAGCCCAACTTCTTGGGCATGTG 5700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5641 ATACGATATGATCCAGAAGAGCTTCCATTTCGCCCTGCAGCCCAACTTCTTGGGCATGTG 5700

Qy       5701 CACTAACTACAAAGAAAGGCTCTGTTACATCAACAATAGTGTGTCTAACAAGCCGGCCAT 5760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5701 CACTAACTACAAAGAAAGGCTCTGTTACATCAACAATAGTGTGTCTAACAAGCCGGCCAT 5760

Qy       5761 CATTCTTAGTTCACTGGTGGGAAACCTCGTCGATCAGAGCAAGCAAGGTATTGTCTTTAA 5820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5761 CATTCTTAGTTCACTGGTGGGAAACCTCGTCGATCAGAGCAAGCAAGGTATTGTCTTTAA 5820

Qy       5821 CGAAGCAAGCTGGGCTCAATTGCGTAGGGAACTGCTTGGCGGTGCATTGTCCCTTCCTGA 5880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5821 CGAAGCAAGCTGGGCTCAATTGCGTAGGGAACTGCTTGGCGGTGCATTGTCCCTTCCTGA 5880

Qy       5881 CCCAATGTACAAGAGCGACAGTTGGCTCGGGCGCGGAGAGCCTACCCACATTATTGACTA 5940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   5881 CCCAATGTACAAGAGCGACAGTTGGCTCGGGCGCGGAGAGCCTACCCACATTATTGACTA 5940

Qy   5941 CCTGAAATTCTCCATCGCCAGGCCTGCGATTGACAAGGAACTGGAAGCCTTCCACAATGC 6000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5941 CCTGAAATTCTCCATCGCCAGGCCTGCGATTGACAAGGAACTGGAAGCCTTCCACAATGC 6000

Qy   6001 CATGAAAGCGGCCAAGGATACAGAAGACGGCGCTCACTTTTGGGATCCGGATCTCGCTTC 6060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6001 CATGAAAGCGGCCAAGGATACAGAAGACGGCGCTCACTTTTGGGATCCGGATCTCGCTTC 6060

Qy   6061 CTACTACACGTTCTTCAAGGAGATTAGCGACAAGTCGCGATCGTCCGCACTGCTATTCAC 6120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6061 CTACTACACGTTCTTCAAGGAGATTAGCGACAAGTCGCGATCGTCCGCACTGCTATTCAC 6120

Qy   6121 GACTCTGAAGAACCGTATCGGCGAAGTCGAGAAAGAATATGGCAGGTTGGTCAAAAACAA 6180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6121 GACTCTGAAGAACCGTATCGGCGAAGTCGAGAAAGAATATGGCAGGTTGGTCAAAAACAA 6180

Qy   6181 GGAGATGAGAGACAGCAAGGACCCCTACCCTGTCCGCGTCAACCAGGTTTATGAAAAATG 6240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6181 GGAGATGAGAGACAGCAAGGACCCCTACCCTGTCCGCGTCAACCAGGTTTATGAAAAATG 6240

Qy   6241 GTGCGCCATCACGCCTGAGGCGATGGACAAATCCGGAGCAAATTATGATTCTAAGGTGAT 6300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6241 GTGCGCCATCACGCCTGAGGCGATGGACAAATCCGGAGCAAATTATGATTCTAAGGTGAT 6300

Qy   6301 CAGGTTGCTGGAGCTGTCCTTCCTCGCGGACCGTGAGATGAATACATGGGCATTGCTGAG 6360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6301 CAGGTTGCTGGAGCTGTCCTTCCTCGCGGACCGTGAGATGAATACATGGGCATTGCTGAG 6360

Qy   6361 GGCTAGCACGGCTTTCAAGCTGTACTACCACAAGAGCCCCAAGTTCGTGTGGCAGATGGC 6420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6361 GGCTAGCACGGCTTTCAAGCTGTACTACCACAAGAGCCCCAAGTTCGTGTGGCAGATGGC 6420

Qy   6421 GGGCAGACAGCTCGCGTACATTAAGGCGCAGATGACGAGCAGACCCGGTGAAGGCGCCCC 6480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6421 GGGCAGACAGCTCGCGTACATTAAGGCGCAGATGACGAGCAGACCCGGTGAAGGCGCCCC 6480

Qy   6481 GGCGTTGATGACCGCGTTCATGTATGCGGGCTTGATGCCGGATAAGAAGTTTACGAAGCA 6540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6481 GGCGTTGATGACCGCGTTCATGTATGCGGGCTTGATGCCGGATAAGAAGTTTACGAAGCA 6540

Qy   6541 GTATGTGGCCAGGCTGGAGGGCGATGGATCGGAGTACCCTGATCCGGAGGTCTATGAAGT 6600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6541 GTATGTGGCCAGGCTGGAGGGCGATGGATCGGAGTACCCTGATCCGGAGGTCTATGAAGT 6600

Qy   6601 GCTGGGCGATGATGATTTTGATGGAATTGGTTTCACAGGGAATGGCGATTATTAGTCGGG 6660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6601 GCTGGGCGATGATGATTTTGATGGAATTGGTTTCACAGGGAATGGCGATTATTAGTCGGG 6660

Qy   6661 TTGCTTGGAGGGGGCCGCTGCATCCTTATTTTTCTGATTTTCCGAGTCTTTGGTGCTGGG 6720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6661 TTGCTTGGAGGGGGCCGCTGCATCCTTATTTTTCTGATTTTCCGAGTCTTTGGTGCTGGG 6720

Qy   6721 AATGGGAATCTTCTGACCATGTTCCCTATGGGTGAGTGCTCATTTTATGGACTTCAGTTT 6780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6721 AATGGGAATCTTCTGACCATGTTCCCTATGGGTGAGTGCTCATTTTATGGACTTCAGTTT 6780

Qy   6781 TGGAGTTTTCTTTCGGCTTTGAGATCTGCCTGAGCAAGCGAGGCGTGGGACATTCGGTTG 6840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6781 TGGAGTTTTCTTTCGGCTTTGAGATCTGCCTGAGCAAGCGAGGCGTGGGACATTCGGTTG 6840

Qy   6841 GTCGCTAGAACATAAAACGAAAATGGGAGATGATATGATTCTTCGTGGCGTGATAATATT 6900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6841 GTCGCTAGAACATAAAACGAAAATGGGAGATGATATGATTCTTCGTGGCGTGATAATATT 6900

Qy   6901 GAGATATTTGGGTGCTGTGCTTTGTATCCATGTCGCAGATATTCGTTCCTACTTGTTGCC 6960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6901 GAGATATTTGGGTGCTGTGCTTTGTATCCATGTCGCAGATATTCGTTCCTACTTGTTGCC 6960

Qy   6961 CGCGAAGCAGATTTGGTCCAACTTTCAAACAATAGTCAGCAACCTTGAGGAAACTATCGA 7020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6961 CGCGAAGCAGATTTGGTCCAACTTTCAAACAATAGTCAGCAACCTTGAGGAAACTATCGA 7020

Qy   7021 TCTCGATACAGACGCTTAGAATAAATAGCGTCCTCGCTCCCTTAGCCTCCGACCTCCATC 7080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7021 TCTCGATACAGACGCTTAGAATAAATAGCGTCCTCGCTCCCTTAGCCTCCGACCTCCATC 7080

Qy   7081 CTAGGTGGAGCACCCGGGATCTCCTAGACTCCTTTCAGACCTGTCTAAGTGAGATATATA 7140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7081 CTAGGTGGAGCACCCGGGATCTCCTAGACTCCTTTCAGACCTGTCTAAGTGAGATATATA 7140

Qy   7141 GTGCAAAGCAGTTGGTGTTACACTTTGGTATTACATCAACAAGACAATTTGAGATAGACT 7200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7141 GTGCAAAGCAGTTGGTGTTACACTTTGGTATTACATCAACAAGACAATTTGAGATAGACT 7200

Qy   7201 CATTCCCTCCGGTTTTCTTCCGATTGCGACTCTATTTGATGATTTTTTTTCCATCTACCC 7260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7201 CATTCCCTCCGGTTTTCTTCCGATTGCGACTCTATTTGATGATTTTTTTTCCATCTACCC 7260

Qy   7261 CTGACAAGCATAACTCTAGCAAATCTTTTTGCTGAATGAATGTAGGCTCTTTGGGATTTC 7320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7261 CTGACAAGCATAACTCTAGCAAATCTTTTTGCTGAATGAATGTAGGCTCTTTGGGATTTC 7320

Qy   7321 GAGGTTTTGACCTTGAAACCTTTGATCTTTTGTTCTTTTGCCTTTTCGTTGAGTTGGAAA 7380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7321 GAGGTTTTGACCTTGAAACCTTTGATCTTTTGTTCTTTTGCCTTTTCGTTGAGTTGGAAA 7380

Qy   7381 GTATATCAACTCTACCTGATCACATCGGAAGCAACAACAACAACAACAACAACAACAACA 7440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7381 GTATATCAACTCTACCTGATCACATCGGAAGCAACAACAACAACAACAACAACAACAACA 7440

Qy   7441 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7441 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7500

Qy   7501 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7501 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7560

Qy   7561 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7561 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7620

Qy   7621 ACAACAACAACAACAACAACAACAACAGCAACAACAACAACAACAACAACAACAACAACA 7680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7621 ACAACAACAACAACAACAACAACAACAGCAACAACAACAACAACAACAACAACAACAACA 7680

Qy   7681 ACAACAACAAGACAACAACAAGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAG 7740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7681 ACAACAACAAGACAACAACAAGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAG 7740

Qy   7741 ACCGAGCGACCTCATAGCTTATGGGAAAATAATACACGTCCAGAGATACATAACAAATAC 7800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7741 ACCGAGCGACCTCATAGCTTATGGGAAAATAATACACGTCCAGAGATACATAACAAATAC 7800

Qy   7801 AGCAACAACAATGAAGTTCAAAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCAC 7860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7801 AGCAACAACAATGAAGTTCAAAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCAC 7860

Qy   7861 CACCACCCACAGCATAAACCTCGAGTGTATGAGTGACGAGTACAACATGAACCGTCGACG 7920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7861 CACCACCCACAGCATAAACCTCGAGTGTATGAGTGACGAGTACAACATGAACCGTCGACG 7920

Qy   7921 ATCCCTATAGTGAGTCGTATTATGCGGCCGCGAATTCGATATCAAGCTTATCGATACCGT 7980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7921 ATCCCTATAGTGAGTCGTATTATGCGGCCGCGAATTCGATATCAAGCTTATCGATACCGT 7980

Qy   7981 CGACCTCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACAATCCCA 8040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7981 CGACCTCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACAATCCCA 8040

Qy   8041 GTCAC 8045
          |||||
Db   8041 GTCAC 8045
```

```
RESULT 2
NCR133528
LOCUS       NCR133528               4206 bp    DNA     linear   PLN 10-MAY-1999
DEFINITION  Neurospora crassa qde-1 gene, partial.
ACCESSION   AJ133528
VERSION     AJ133528.1  GI:4803726
KEYWORDS    qde-1 gene; RNA-dependent RNA polymerase.
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Cogoni,C. and Macino,G.
  TITLE     Gene silencing in neurospora crassa requires a protein homologous
            to RNA-dependent RNA polymerase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4206)
  AUTHORS   Cogoni,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-1999) Cogoni C., Dept. Biotecnologie Cellulari ed
            Ematologia,, Universita di Roma La Sapienza, Viale Regina Elena
            324, 00161 Rome, ITALY
FEATURES             Location/Qualifiers
     source          1..4206
                     /organism="Neurospora crassa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5141"
     gene            1..4206
                     /gene="qde-1"
     CDS             1..>4206
                     /gene="qde-1"
                     /function="gene silencing"
                     /note="putative"
                     /codon_start=1
                     /product="RNA-dependent RNA polymerase"
                     /protein_id="CAB42634.1"
                     /db_xref="GI:4803727"
                     /db_xref="GOA:Q9Y7G6"
                     /db_xref="SPTREMBL:Q9Y7G6"
                     /translation="MNPITPRKRNSPVEEIINRLNNDYNLGLQCVADTTLTPHRRKEL
                     AESDEDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLP
                     WSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNT
                     KRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQC
                     TKRAKGKLSDNVAAAAAPPVPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQ
                     GTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKS
                     HIDNQPLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAP
                     VVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDV
                     TDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTA
                     PLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQ
                     WLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFRP
                     DVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKTYAI
                     MTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFG
                     SAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQL
                     LPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRV
                     SHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRV
                     GRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQ
                     RVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEM
                     PLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNY
                     KERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDP
                     MYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWDPDLA
                     SYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQVY
                     EKWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHKSPKF
                     VWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGDGSEYP
                     DPEVYEVLGDDDFDGIGFTGNGDY"
ORIGIN

  Query Match             52.3%;  Score 4206;  DB 8;  Length 4206;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2447 ATGAACCCTATTACTCCTAGGAAGAGGAATAGCCCCGTCGAGGAAATCATAAACCGGCTC 2506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAACCCTATTACTCCTAGGAAGAGGAATAGCCCCGTCGAGGAAATCATAAACCGGCTC 60

Qy       2507 AATAACGACTACAACCTGGGCCTCCAGTGTGTCGCAGACACAACTCTCACCCCCCACCGC 2566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AATAACGACTACAACCTGGGCCTCCAGTGTGTCGCAGACACAACTCTCACCCCCCACCGC 120

Qy       2567 CGGAAGGAGCTGGCCGAGAGTGACGAGGATTTCGGTCGCCATGACAAGATCTACAGAGCC 2626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGGAAGGAGCTGGCCGAGAGTGACGAGGATTTCGGTCGCCATGACAAGATCTACAGAGCC 180

Qy       2627 CTGAACTTTCTCTACTGGCGGAAGGATGACTCCCTGAACCAGGCAGAAGCCAACTTCTTC 2686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGAACTTTCTCTACTGGCGGAAGGATGACTCCCTGAACCAGGCAGAAGCCAACTTCTTC 240

Qy       2687 ATCGAGGCCAAAGCTGCGAGCTCGAACTGGGTGCCCAAAGCCCACGCCGACCCTGACACG 2746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATCGAGGCCAAAGCTGCGAGCTCGAACTGGGTGCCCAAAGCCCACGCCGACCCTGACACG 300

Qy       2747 CTTCCGTGGTCCAAGGAACCTCCCCGCGCCGCTACTGCCGGCCAACAATGGGCATTGCAG 2806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTCCGTGGTCCAAGGAACCTCCCCGCGCCGCTACTGCCGGCCAACAATGGGCATTGCAG 360

Qy       2807 ACTGTGTTGCTCGAGGTGCTTAATAGGTTTATGCCACCTCCCAATAACACACCAGGTCGA 2866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTGTGTTGCTCGAGGTGCTTAATAGGTTTATGCCACCTCCCAATAACACACCAGGTCGA 420

Qy       2867 ACGTTTGGCAGAACTCTAAGCGGCCCAAGTGGCCTGAGCCGCCCAACCTCTACCAACACC 2926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGTTTGGCAGAACTCTAAGCGGCCCAAGTGGCCTGAGCCGCCCAACCTCTACCAACACC 480

Qy       2927 AAACGCAAGGATGAGCCCGCCAATGTCACTTTCGCTGATCCGCCCAAACGCTCGTTGACT 2986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACGCAAGGATGAGCCCGCCAATGTCACTTTCGCTGATCCGCCCAAACGCTCGTTGACT 540

Qy       2987 CGCTCTGCCACAGGTCCTCCTATTCACGGCGCGGCGATACCCCTAAAGTTCCCCGATCCA 3046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGCTCTGCCACAGGTCCTCCTATTCACGGCGCGGCGATACCCCTAAAGTTCCCCGATCCA 600

Qy       3047 GTGAATACCGGTTCCAAACGACCATCTCTCGAGAGTGAGAATCTCAATCAGTGCACCAAG 3106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTGAATACCGGTTCCAAACGACCATCTCTCGAGAGTGAGAATCTCAATCAGTGCACCAAG 660

Qy       3107 CGGGCCAAGGGCAAGCTGTCTGATAATGTTGCCGCTGCCGCCGCCCCGCCCGTGCCTATT 3166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGGGCCAAGGGCAAGCTGTCTGATAATGTTGCCGCTGCCGCCGCCCCGCCCGTGCCTATT 720

Qy       3167 GCGAGCGCTTTGGACAAGGTACCGACTCGAAGGCATGCCAATACGAGAGATCCCACGGCG 3226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCGAGCGCTTTGGACAAGGTACCGACTCGAAGGCATGCCAATACGAGAGATCCCACGGCG 780

Qy       3227 ACAGGTCATAGACGAGCGGACCAGGTGGATTCCTTTGATACATCTCAAGGCACTTCCTAT 3286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACAGGTCATAGACGAGCGGACCAGGTGGATTCCTTTGATACATCTCAAGGCACTTCCTAT 840

Qy       3287 GGTTCGAGTGTCTTCAGCGCTTGCCGTCACAATCAGAGCACTACCCAGAGTAGTTTTGAG 3346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGTTCGAGTGTCTTCAGCGCTTGCCGTCACAATCAGAGCACTACCCAGAGTAGTTTTGAG 900

Qy       3347 GCTCCTCCTTCACAGCCCAGAGAGAAGCGGCCTGTGGATGCCACGGTCTTTGAGGCTGGA 3406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCTCCTCCTTCACAGCCCAGAGAGAAGCGGCCTGTGGATGCCACGGTCTTTGAGGCTGGA 960

Qy       3407 CACTTGATTGAGTCTCCTAGCAAAGGAAGAACAACCAAGTCCCACATAGATAACCAGCCC 3466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CACTTGATTGAGTCTCCTAGCAAAGGAAGAACAACCAAGTCCCACATAGATAACCAGCCC 1020

Qy       3467 CTTTCATCGTCTTCCCAGGGTGAAACTTCGTTCAGCACTTACTATGAGTCGTTTCCAAGT 3526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTTTCATCGTCTTCCCAGGGTGAAACTTCGTTCAGCACTTACTATGAGTCGTTTCCAAGT 1080

Qy       3527 TCCGGCGGCGAGGGCGCAATTCCCGAGCCGAGTCGCTCAAATGGACTGGCTCGGAGCGAA 3586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TCCGGCGGCGAGGGCGCAATTCCCGAGCCGAGTCGCTCAAATGGACTGGCTCGGAGCGAA 1140
```

```
Qy    3587 GAAAGCGCTCGATCTCAGGTTCAAGTTCATGCCCCGGTGGTTGCAGCTCGGCTGAGAAAT 3646
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GAAAGCGCTCGATCTCAGGTTCAAGTTCATGCCCCGGTGGTTGCAGCTCGGCTGAGAAAT 1200

Qy    3647 ATTTGGCCGAAATTTCCCAAATGGCTACACGAAGCTCCTCTCGCTGTTGCATGGGAAGTT 3706
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ATTTGGCCGAAATTTCCCAAATGGCTACACGAAGCTCCTCTCGCTGTTGCATGGGAAGTT 1260

Qy    3707 ACCAGACTCTTTATGCACTGCAAAGTAGACTTGGAAGACGAGAGCCTGGGCCTAAAGTAC 3766
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ACCAGACTCTTTATGCACTGCAAAGTAGACTTGGAAGACGAGAGCCTGGGCCTAAAGTAC 1320

Qy    3767 GACCCTTCCTGGTCTACCGCGCGCGATGTCACAGATATCTGGAAGACTCTCTACCGGCTT 3826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 GACCCTTCCTGGTCTACCGCGCGCGATGTCACAGATATCTGGAAGACTCTCTACCGGCTT 1380

Qy    3827 GATGCTTTCCGTGGTAAACCCTTTCCAGAAAAGCCGCCCAACGACGTGTTCGTGACCGCA 3886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 GATGCTTTCCGTGGTAAACCCTTTCCAGAAAAGCCGCCCAACGACGTGTTCGTGACCGCA 1440

Qy    3887 ATGACGGGCAACTTTGAGAGCAAAGGTAGTGCCGTTGTTCTCTCTGCTGTTCTAGACTAC 3946
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 ATGACGGGCAACTTTGAGAGCAAAGGTAGTGCCGTTGTTCTCTCTGCTGTTCTAGACTAC 1500

Qy    3947 AATCCGGACAACTCGCCTACTGCGCCCCTTTACCTTGTGAAGCTGAAGCCGCTCATGTTC 4006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 AATCCGGACAACTCGCCTACTGCGCCCCTTTACCTTGTGAAGCTGAAGCCGCTCATGTTC 1560

Qy    4007 GAGCAGGGCTGTCGACTCACCCGTCGGTTCGGTCCTGATAGGTTTTTCGAGATCCTTATA 4066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 GAGCAGGGCTGTCGACTCACCCGTCGGTTCGGTCCTGATAGGTTTTTCGAGATCCTTATA 1620

Qy    4067 CCGTCGCCTACGAGCACCAGCCCAAGTGTACCGCCGGTGGTCAGCAAACAACCAGGTGCG 4126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 CCGTCGCCTACGAGCACCAGCCCAAGTGTACCGCCGGTGGTCAGCAAACAACCAGGTGCG 1680

Qy    4127 GTCGAAGAAGTCATCCAGTGGCTCACGATGGGGCAACATTCTCTTGTAGGCCGCCAATGG 4186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 GTCGAAGAAGTCATCCAGTGGCTCACGATGGGGCAACATTCTCTTGTAGGCCGCCAATGG 1740

Qy    4187 CGCGCTTTCTTCGCCAAAGATGCCGGATACAGGAAACCTCTCAGGGAGTTCCAGCTCCGC 4246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 CGCGCTTTCTTCGCCAAAGATGCCGGATACAGGAAACCTCTCAGGGAGTTCCAGCTCCGC 1800

Qy    4247 GCCGAGGACCCGAAACCCATCATCAAGGAGAGAGTCCACTTCTTTGCCGAGACCGGCATT 4306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 GCCGAGGACCCGAAACCCATCATCAAGGAGAGAGTCCACTTCTTTGCCGAGACCGGCATT 1860

Qy    4307 ACGTTCCGACCTGATGTGTTCAAGACGAGATCTGTCGTTCCGGCAGAGGAACCTGTAGAG 4366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 ACGTTCCGACCTGATGTGTTCAAGACGAGATCTGTCGTTCCGGCAGAGGAACCTGTAGAG 1920

Qy    4367 CAACGGACCGAGTTCAAAGTTAGTCAAATGCTGGACTGGCTCCTGCAACTCGACAACAAC 4426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 CAACGGACCGAGTTCAAAGTTAGTCAAATGCTGGACTGGCTCCTGCAACTCGACAACAAC 1980

Qy    4427 ACTTGGCAGCCGCACCTCAAGTTGTTCTCCCGTATCCAGCTCGGTCTGAGTAAGACATAT 4486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 ACTTGGCAGCCGCACCTCAAGTTGTTCTCCCGTATCCAGCTCGGTCTGAGTAAGACATAT 2040

Qy    4487 GCCATTATGACATTGGAGCCTCACCAGATCAGACACCACAAGACCGATCTTCTTTCACCT 4546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 GCCATTATGACATTGGAGCCTCACCAGATCAGACACCACAAGACCGATCTTCTTTCACCT 2100

Qy    4547 TCAGGCACTGGCGAAGTGATGAATGACGGTGTAGGCCGCATGTCGCGAAGCGTGGCCAAG 4606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 TCAGGCACTGGCGAAGTGATGAATGACGGTGTAGGCCGCATGTCGCGAAGCGTGGCCAAG 2160

Qy    4607 AGGATACGCGATGTTCTCGGTTTGGGTGATGTGCCCTCTGCTGTGCAAGGGCGGTTTGGT 4666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 AGGATACGCGATGTTCTCGGTTTGGGTGATGTGCCCTCTGCTGTGCAAGGGCGGTTTGGT 2220

Qy    4667 TCGGCCAAGGGAATGTGGGTTATCGACGTTGACGACACAGGCGATGAGGATTGGATCGAG 4726
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 TCGGCCAAGGGAATGTGGGTTATCGACGTTGACGACACAGGCGATGAGGATTGGATCGAG 2280

Qy    4727 ACATACCCGTCCCAGCGCAAGTGGGAATGCGACTTCGTTGATAAACATCAACGTACCCTC 4786
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 ACATACCCGTCCCAGCGCAAGTGGGAATGCGACTTCGTTGATAAACATCAACGTACCCTC 2340

Qy    4787 GAAGTCCGGAGCGTGGCTTCTGAACTGAAGTCAGCTGGTCTCAACCTACAGCTGTTACCT 4846
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 GAAGTCCGGAGCGTGGCTTCTGAACTGAAGTCAGCTGGTCTCAACCTACAGCTGTTACCT 2400

Qy    4847 GTCCTGGAAGATAGAGCCAGGGACAAGGTGAAGATGCGCCAGGCAATCGGTGACCGTCTT 4906
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 GTCCTGGAAGATAGAGCCAGGGACAAGGTGAAGATGCGCCAGGCAATCGGTGACCGTCTT 2460

Qy    4907 ATCAACGATTTGCAACGACAGTTCAGCGAGCAAAAGCATGCTTTGAATCGCCCAGTGGAA 4966
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 ATCAACGATTTGCAACGACAGTTCAGCGAGCAAAAGCATGCTTTGAATCGCCCAGTGGAA 2520

Qy    4967 TTTCGCCAATGGGTTTACGAGAGTTATTCCAGTCGCGCAACTCGAGTCAGCCACGGCCGT 5026
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 TTTCGCCAATGGGTTTACGAGAGTTATTCCAGTCGCGCAACTCGAGTCAGCCACGGCCGT 2580

Qy    5027 GTGCCTTTTCTTGCTGGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAAC 5086
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 GTGCCTTTTCTTGCTGGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAAC 2640

Qy    5087 AGTGGGTTCGATCCCAAGAAGCAAAAGTACTTGCAAGACATCGCCTGGGATCTTCAAAAG 5146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 AGTGGGTTCGATCCCAAGAAGCAAAAGTACTTGCAAGACATCGCCTGGGATCTTCAAAAG 2700

Qy    5147 CGGAAATGTGACACGTTGAAGTCCAAGCTGAACATCCGTGTCGGTCGATCAGCATACATT 5206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 CGGAAATGTGACACGTTGAAGTCCAAGCTGAACATCCGTGTCGGTCGATCAGCATACATT 2760

Qy    5207 TACATGATTGCCGATTTCTGGGGTGTGCTTGAGGAAAATGAGGTTCATGTCGGATTCTCC 5266
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 TACATGATTGCCGATTTCTGGGGTGTGCTTGAGGAAAATGAGGTTCATGTCGGATTCTCC 2820

Qy    5267 TCAAAGTTCAGAGACGAGGAGGAGTCTTTTACACTCCTATCGGACTGTGATGTCCTCGTG 5326
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 TCAAAGTTCAGAGACGAGGAGGAGTCTTTTACACTCCTATCGGACTGTGATGTCCTCGTG 2880

Qy    5327 GCGCGATCCCCAGCCCATTTCCCTAGTGATATCCAACGGGTTCGAGCAGTCTTCAAGCCA 5386
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2881 GCGCGATCCCCAGCCCATTTCCCTAGTGATATCCAACGGGTTCGAGCAGTCTTCAAGCCA 2940

Qy    5387 GAGCTCCACAGTCTCAAGGATGTAATCATCTTCTCTACTAAAGGAGATGTACCGCTTGCT 5446
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2941 GAGCTCCACAGTCTCAAGGATGTAATCATCTTCTCTACTAAAGGAGATGTACCGCTTGCT 3000

Qy    5447 AAGAAGCTATCTGGTGGAGACTACGACGGCGATATGGCCTGGGTCTGCTGGGATCCGGAG 5506
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3001 AAGAAGCTATCTGGTGGAGACTACGACGGCGATATGGCCTGGGTCTGCTGGGATCCGGAG 3060

Qy    5507 ATCGTCGATGGTTTCGTCAATGCGGAAATGCCTCTGGAGCCCGACCTGTCTAGGTACCTA 5566
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3061 ATCGTCGATGGTTTCGTCAATGCGGAAATGCCTCTGGAGCCCGACCTGTCTAGGTACCTA 3120

Qy    5567 AAGAAGGACAAAACGACTTTCAAACAACTTATGGCCTCACACGGCACGGGCTCAGCGGCC 5626
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3121 AAGAAGGACAAAACGACTTTCAAACAACTTATGGCCTCACACGGCACGGGCTCAGCGGCC 3180

Qy    5627 AAAGAGCAGACTACATACGATATGATCCAGAAGAGCTTCCATTTCGCCCTGCAGCCCAAC 5686
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3181 AAAGAGCAGACTACATACGATATGATCCAGAAGAGCTTCCATTTCGCCCTGCAGCCCAAC 3240

Qy    5687 TTCTTGGGCATGTGCACTAACTACAAAGAAAGGCTCTGTTACATCAACAATAGTGTGTCT 5746
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3241 TTCTTGGGCATGTGCACTAACTACAAAGAAAGGCTCTGTTACATCAACAATAGTGTGTCT 3300

Qy    5747 AACAAGCCGGCCATCATTCTTAGTTCACTGGTGGGAAACCTCGTCGATCAGAGCAAGCAA 5806
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3301 AACAAGCCGGCCATCATTCTTAGTTCACTGGTGGGAAACCTCGTCGATCAGAGCAAGCAA 3360

Qy        5807 GGTATTGTCTTTAACGAAGCAAGCTGGGCTCAATTGCGTAGGGAACTGCTTGGCGGTGCA 5866
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3361 GGTATTGTCTTTAACGAAGCAAGCTGGGCTCAATTGCGTAGGGAACTGCTTGGCGGTGCA 3420

Qy        5867 TTGTCCCTTCCTGACCCAATGTACAAGAGCGACAGTTGGCTCGGGCGCGGAGAGCCTACC 5926
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3421 TTGTCCCTTCCTGACCCAATGTACAAGAGCGACAGTTGGCTCGGGCGCGGAGAGCCTACC 3480

Qy        5927 CACATTATTGACTACCTGAAATTCTCCATCGCCAGGCCTGCGATTGACAAGGAACTGGAA 5986
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3481 CACATTATTGACTACCTGAAATTCTCCATCGCCAGGCCTGCGATTGACAAGGAACTGGAA 3540

Qy        5987 GCCTTCCACAATGCCATGAAAGCGGCCAAGGATACAGAAGACGGCGCTCACTTTTGGGAT 6046
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3541 GCCTTCCACAATGCCATGAAAGCGGCCAAGGATACAGAAGACGGCGCTCACTTTTGGGAT 3600

Qy        6047 CCGGATCTCGCTTCCTACTACACGTTCTTCAAGGAGATTAGCGACAAGTCGCGATCGTCC 6106
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3601 CCGGATCTCGCTTCCTACTACACGTTCTTCAAGGAGATTAGCGACAAGTCGCGATCGTCC 3660

Qy        6107 GCACTGCTATTCACGACTCTGAAGAACCGTATCGGCGAAGTCGAGAAAGAATATGGCAGG 6166
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3661 GCACTGCTATTCACGACTCTGAAGAACCGTATCGGCGAAGTCGAGAAAGAATATGGCAGG 3720

Qy        6167 TTGGTCAAAAACAAGGAGATGAGAGACAGCAAGGACCCCTACCCTGTCCGCGTCAACCAG 6226
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3721 TTGGTCAAAAACAAGGAGATGAGAGACAGCAAGGACCCCTACCCTGTCCGCGTCAACCAG 3780

Qy        6227 GTTTATGAAAAATGGTGCGCCATCACGCCTGAGGCGATGGACAAATCCGGAGCAAATTAT 6286
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3781 GTTTATGAAAAATGGTGCGCCATCACGCCTGAGGCGATGGACAAATCCGGAGCAAATTAT 3840

Qy        6287 GATTCTAAGGTGATCAGGTTGCTGGAGCTGTCCTTCCTCGCGGACCGTGAGATGAATACA 6346
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3841 GATTCTAAGGTGATCAGGTTGCTGGAGCTGTCCTTCCTCGCGGACCGTGAGATGAATACA 3900

Qy        6347 TGGGCATTGCTGAGGGCTAGCACGGCTTTCAAGCTGTACTACCACAAGAGCCCCAAGTTC 6406
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3901 TGGGCATTGCTGAGGGCTAGCACGGCTTTCAAGCTGTACTACCACAAGAGCCCCAAGTTC 3960

Qy        6407 GTGTGGCAGATGGCGGGCAGACAGCTCGCGTACATTAAGGCGCAGATGACGAGCAGACCC 6466
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3961 GTGTGGCAGATGGCGGGCAGACAGCTCGCGTACATTAAGGCGCAGATGACGAGCAGACCC 4020

Qy        6467 GGTGAAGGCGCCCCGGCGTTGATGACCGCGTTCATGTATGCGGGCTTGATGCCGGATAAG 6526
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4021 GGTGAAGGCGCCCCGGCGTTGATGACCGCGTTCATGTATGCGGGCTTGATGCCGGATAAG 4080

Qy        6527 AAGTTTACGAAGCAGTATGTGGCCAGGCTGGAGGGCGATGGATCGGAGTACCCTGATCCG 6586
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4081 AAGTTTACGAAGCAGTATGTGGCCAGGCTGGAGGGCGATGGATCGGAGTACCCTGATCCG 4140

Qy        6587 GAGGTCTATGAAGTGCTGGGCGATGATGATTTTGATGGAATTGGTTTCACAGGGAATGGC 6646
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4141 GAGGTCTATGAAGTGCTGGGCGATGATGATTTTGATGGAATTGGTTTCACAGGGAATGGC 4200

Qy        6647 GATTAT 6652
               ||||||
Db        4201 GATTAT 4206
```

```
RESULT 3
AC116986
LOCUS       AC116986               333321 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 2234041-2567370 strain
            AX4, complete sequence.
ACCESSION   AC116986 AC116425
VERSION     AC116986.2  GI:28829247
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 333321)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 333321)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 333321)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 333321)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 4, 2003 this sequence version replaced
            gi:19774405, gi:20042949.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the University Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .333321
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="2234041-2567370"
     CDS             complement(join(2281. .4838,5022. .5160))
                     /note="GeneID exon scores (in order of location ranges):
                     218.28, 7.77 - GSCJ_ID dd_03279"
                     /codon_start=1
                     /product="similar to Plasmodium falciparum (isolate 3D7).
                     Hypothetical protein"
                     /protein_id="AAM08764.2"
                     /db_xref="GI:28829248"
                     /translation="MKKLFDIISPSKNIQQQSASGKESSSSSSSSSSPSPSKNRSCFN
                     GIGINSKINNINIQEDSISNGNNNEKYPQIERDNIIKIVLLGGYQVGKSSIIRRYLQG
                     CFTTSYIPTIGIDLSTLLLKRKQSQQQQQQQQQQQQQQQQQQQQQQPSPQSSQQPPPQ
                     QQQQQQQQQQQQQQQTPQRNKNENYDENDYLSDYSDSESKDDSSLNQYLTQSPNQYIV
                     QFWDVAYAEIKGKHFSSILSGASAIIMVYDSVNASSILAIDEWKTILNTSPFIQINNG
                     NNNNNNNNNNNNNSNNNSNNNNSNNNSFNSNNYKKIPLALFANKSDIGEPVISPTDLD
                     YYCNACGYSIWRFTSCKKDLNIISGIDSLISLVIENKNQELSDIRNRLLNEKKEKRKL
                     IYQLKKNVNLNTSVKYTRSINKSSSGNKNQDNNNVDDIDIVKNLTKELVLSQSCPNIS
                     NYGKDVNNDDDDDNNNTIENNENENNENVKKQQEKEKEQKEKEKEKLQQFKHQSEIEK
                     VNELLEMTNEHYSKFQKSLEEIMSKSNGTTRYNDLVQLEKQRSNEYVKICNNLKKLIQ
                     PSAFRQHHHHHHHHHHHKPSQKITSDLLSKILDENILKWNTLISNLQLEISLKLNNQN
                     NVYQSNNNNNNNNVKQNNKVAQSYNAVSLLIKDDQEYRYNNNNNNNNGGRKNSFSNSS
                     NNINNNNNNNINNNNNNNNNNNNNNNNNNNNNNNNNNSSINNSIKLNINNNPKYKMG
                     FGFKHGGKNLSSNFSINDHSIDTDNPSMTSSSITSNSSMNSFSSSPSSSPPNDDINDN
                     RHTNNINRVATNGDGNVDGGGGDNIGDNDSSSLNSSSYQNHHNHHNHHRHNFNNDSNN
                     NSYSPSTSHNGIGGDDFLLPSPSLLPSSYNLTVTVLPDLKYK"
     CDS             join(5597. .5760,5890. .6593,6692. .10869)
                     /note="GeneID exon scores (in order of location ranges):
                     7.48, 52.43, 406.35 - GSCJ_ID dd_01879"
                     /codon_start=1
```

```
                     /product="similar to Arabidopsis thaliana (Mouse-ear
                     cress). Putative UDP-glucose:glycoprotein
                     glucosyltransferase, 101200- 91134"
                     /protein_id="AAM08766.2"
                     /db_xref="GI:28829249"
                     /translation="MARIFKFFVFLLIVFISNVLLLVESNEGDNSFSSKSIQLSLVSN
                     WGETPSYLEAAEFLHNQDKSLFWKFIEEFNKIDFSTNYSDKIYYESTISLMKSVLSSN
                     TQFLSEFLSIDLAMRTYSPRVETYRQLAISNMKLNNIEHSITTADNKTITLFNSGGWV
                     QIKNKIITDVNEINESLFKDVAVVDDEENEFIRLYDFDHIFPTLANTVSSSSSSPSSI
                     PIVILYVDIKSEFFKLVHPKLKQFSQMGKIKYCLRYVVQESNQKLNLQGYGYELSIKN
                     LEYKVMDDSAIKKDIIIDGVKSKTIINIPNEDVQGFNFHKLQKRKPELTSKLSTFRSY
                     LMAKSQEAKELKVWELKDLGIQSAQKIIQSGDPLRSLEYISQKFPTLSNSLSKITLNE
                     SLKSVIESNQKIIPSTTDQTLLLNGRLIDTNELSPIELSRIILEEYEHSTTIQQQGPL
                     SSKTVQDIISAQLPIRIQLLPTKEELELNGGNEPFVSLNNLELDYIYRQWEPKLQSSV
                     LDKPVTSPQDIFIRKNLLTTVIVLDWNNINTFEIIPEIQEMVQGNSLIPTRIQLLFNT
                     KSNNNNNNNNNNNDQNSQTSNFIQGKDLAKVFLTIKNSNLGNRGAFFFITALNYFKKM
                     YIPNELGITRSVLSSSFQAVLQQMGGSVRSLQHALTNTDFDNLLESSNQLIERLELLD
                     TTTSQSTTTTTTTKILPKVFVNGVQVKYSNIDQLSFDLLVSLYDEFDNLKPLFKESIL
                     STTTAQYYETILTSSYWKDNNLPFLKKLNSMISNEKYSHLITNSKNRNQEVDAQNVLK
                     NLLYFRNNENKDEQNLLNLIVIGDFDHYNTRDISLELLRQLEKGELKNCKLTFISNPI
                     DINSVVNTAGNENQILGKLITILKHYGKILTPQLVIGLFEKVQSDPTIIDSFKTMKQI
                     IELSGFDIAANDIWVAQSVNLFKQSSKVCKQYLGIQSTNKSPLSILVNGRIITPPLSY
                     DDAASFIQSDFKLLLEIEMIKAKKTFELLNSDPILKDKSNLKISDLLNKVQSLVGYYY
                     NGNNQLDSNIKRKRIPNSLSISFSHKPPTLSSSSSSSSSNSNDVPLKFLMIINPFNKV
                     SQKLVPMVREFSNKLNIPVDVILNPPVSLSELPLKTYYTYVIKLSSEFNNENVLYNQP
                     LGIATDIPEDRVVTLALDIPSSWLVQPIIAKYDLDNIRLKDLGDEQVLTAVYELENIV
                     IEGSANDMTTDNAPAGLELLLNPISTQTNKTQDTIVMNNFGYYQLKSNPGIWKLTIAP
                     GRSSDIMDMVDHPNQKEKETFVIVPHRLVVIDSLYQSLSSLSVVRKAGQELRPILQPI
                     DEYEKQKEQEKEQKLKQNSSGFFSNLFSSKNDATDSVATHQKKSNLDTIHIFSVASGH
                     LYERFLKIMMLSVVKNTESPIKFWFLKNYLSPAFKEFIPEMAKEYGFQYELVTYKWPW
                     WLRKQTEKQRIIWSYKILFLDVLFPLDVPKIIFVDADQVVRTDLKELWDMDLHGASLG
                     YTPFCDSNKDTEGFRFWKSGYWRQHLAGRSYHISALYVVDLVRFRRLAAGDQLRATYD
                     QLSRDPNSLANLDQDLPNYLQHYVRIHSLPQEWLWCETWCDQESKSKAKTIDLCNNPL
                     TKTPKLENAVRIIDEWTTLDNEAKEFELKIDQSKHHRQIELDHQNQLPNSKPIENIDD
                     ILLNLAESQKDLF"
     CDS             complement(join(10993. .12657,12738. .12853,13014. .13262,
                     13781. .13814,14302. .14529))
                     /note="GeneID exon scores (in order of location ranges):
                     175.98, 4.12, 21.10, -3.61, 20.21 - GSCJ_ID dd_01880"
                     /codon_start=1
                     /product="similar to Neurospora crassa. Probable branching
                     enzyme (bel)"
                     /protein_id="AAM08768.2"
                     /db_xref="GI:28829250"
                     /translation="MNVDDIYADIIIESENDQENDMKQSDENSQSPTLKEDHHNKIKE
                     FEEIIENNLKEISNLKKEIESKDNEIETLSKQIFSIIVKLILRDYSKATDGTKVIHDD
                     PWLEPYKEVIKRRHNQVKNTIQKLEESEGSLLKFSQGYEYFGFNVTKDGVNYREWLPS
                     AHEVYLVGDFNQWNKTSHPLERDNYGRWSIFIPNNSNGECAIPHGSKIKIYLKLANGN
                     FDYRIPAWIKRVEQTKENPVFDGVFWNPSKQYVFKNKSPMKPTELRIYEAHVGMSSEL
                     PEISTYSKFKDTVLPMVKELGYNCIQLMAVMEHAYYASFGYQVTNFFAISSRFGTPEE
                     LKEMIDKAHEMGLLVFLDVVHSHASKNVLDGLNQLDGTDHHYFHSGGRGNHELWDSRL
                     FNYGNWEVMRFLLSNLRFYVDEYHFDGFRFDGVTSMIYTHHGLSPACSYDDYFGGAVD
                     EDALNYLTLANVMLHTLNPSIVTIAEEVTGLATLCRPFSEGGGDFDYRLAMGIPDKWI
                     ELVKEKKDEDWNMGTIAHMLSNRRYKEKNIAYAESHDQSLVGDKTLAFWLMDKEMYTN
                     MSVTTEETPIIDRGMSLHKMIRLITSSLGGDGYLNFMGNEFGHPEWVDFPREGNNNSL
                     HHARRRWDLYRNPLLRYKQLRDFDIAMNKAEQEFRWLSSDFAYISLKHEDDKIIVFER
                     ASLIFIFNFHPSKSFSDYRIGSGVPGKFINVLDSDRKEFGGHVRIGKDNYHYTEDKPW
                     HDRKYSLLIYIPSRTCLVLKKVD"
     CDS             join(15053. .15120,15221. .15743)
                     /note="GeneID exon scores (in order of location ranges):
                     -0.56, 37.71 - GSCJ_ID dd_01882"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAM08769.2"
                     /db_xref="GI:28829251"
                     /translation="MGQKSLGELQYPENMKPLFNVKIEDCKYSYDTKDDTFLSIQYND
                     NNKQAKSNNDSSIGNNKNSKEQSSSTTTTTRRRNVNSKEPIIKTSSNGVTTMGSLESL
                     LNNCSTNDNSSDDDDDDESDTDEEDEDNDNSFENNPNLTEEEKQKIFQEKVKKQQQKK
                     KLKANKKIIRKDPIYWFGISTPTSLKQSQSYFKQGK"
     CDS             complement(join(16609. .16729,16860. .16934,17034. .17080,
                     17169. .17177))
                     /note="GeneID exon scores (in order of location ranges):
                     6.96, 13.18, -2.05, 0.27 - GSCJ_ID dd_01883"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO51794.1"
                     /db_xref="GI:28829252"
                     /translation="MAGPWTIEPGKNHGGIPTWAWPRSTTVHVQVVGGGRGRVRMQAG
                     ASPDEDNDVNGETSFSRSFGGSRLNVTNIGSRTLKVWTA"
     CDS             complement(19502. .21007)
                     /note="GeneID exon scores (in order of location ranges):
                     108.08 - GSCJ_ID dd_01875"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). FutA (Fragment)"
                     /protein_id="AAO51795.1"
                     /db_xref="GI:28829253"
                     /translation="MKSLRNSYSILLKIVIISTCLLLLINIALIYEDPNLIDSNCKTT
                     TTITKTQNNKNNKGLENLNKSKKKLLLILDYKLFNEDTYMKSFSECTSITNQPQCGFT
                     TDISLANESDGILYFIPRTTKKNYNRSYTSVHRTSEKQLFFGWSMESPSMYPLENDDT
                     HLSLNYNVTIGYPLNDYQKYGAGVSDEEYEEKDIQSSHIYTPYGPVPFKGSDHYSYLF
                     KFGKLKEIPKKRDSNIMWMAGNCRNQKLGRTKLVKQMMRFTTIDSYGKCLHNKDLKVN
                     YGPNVLLRSLQKLKILRKYDFVVAIENSRCEDYVTEKLWEPLSVGTIPIYLGATNIDE
                     FLPHPDAIINIANFNSINHLMKYVRDVGLNETLREKHLAWIKKPLPKKFKNLLNQSLN
                     NQNLFCSICKKLVNLDNNDNINYTPLSKPFKQCIEGKFTLPIYTEEEKERGRLLRLQK
                     QQEQEKLELQNNNDNYTNDNEEEEEEEDNDSGNDSGNDSGNDSGNNDSNSI"
     CDS             complement(join(21960. .26819,26946. .27251))
                     /note="GeneID exon scores (in order of location ranges):
                     530.44, 21.78 - GSCJ_ID dd_03196"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Nucleotide exchange factor RasGEF R"
                     /protein_id="AAO51796.1"
                     /db_xref="GI:28829254"
                     /translation="MKEEEIANLGPQEYFNYLEGINKKIIIELKTVANIDTAKDTDNL
                     QNILIESSLALEHNIRNLLYFVDKCPISYDQIVHYHSHFKEVVKIFFQVCQDLIKAAQ
                     ALILNSCDYLTTSNFAKCRGEAIKTIKNTMATCSNFEKLYYNPDLQLEDEVDLVHQQQ
                     QKEQEQQQKEQEQYEQEQLILKQLKEQELKKQQEQQQQQQQQQQQQQQQQQQQQQKHQ
                     EEKEKNDQKEKEEKLKKQLLAEQQKQKQEELERQQQQLQQQQQLLQQQQQLLQQQQQQ
                     QRSDPIIIPPITLKTSQSDLLSGGLSPKVNSPGSPNLGKFTESAERVANFDRGTANII
                     LNMNKIKEISLSGDSEQAPEIVTAARIISENIAIISKELRYKTLGTNLSEHFVSVVQV
                     ARLAIKNKSDQFYQDQLELAIEKFNDTMRKIIYSVKSISKSSMNLRQFGVDESSLGRS
                     PVVSPEKSISPSFTSSTSERILSHEIKNHNNNNNNYNNSSTNNLQTSFSTPSLSSNHS
                     QQPNQQPLQSPLLINQLQSTSSSSSSSSSNLSNSLNSIQLPQATTTTTTATSPSTSTS
                     TSTSTSPNSSSLSISDQDKQLKRKEKQHQIVIPKDVKDSVANQQQQQQQQQNGTTSPR
                     NNESSVTAATTTTTSTTASITTNVNTIPNFPPNKQLPATPTTGTPSTSTPTPQTPTST
                     SQNDKQNENNNKENFVDKQKTLGKLFSKFVHKKRTPLPGFDSSNSSSPSNNSNTTNSS
                     SHSSTNSSPMETSPGVESPKITKSPSQNNILVDSLDIGSDNQQQQQDKLTTTTSTTTI
                     TINNNNNNNNNNNNNNNNNNNNIQQQQQQQQQIPTTPNKLSNSIAMNSSNRLMTPKKE
                     RSFTIGLVSGKHSCVVVETPKSKKLFQHESAKQILSIISMNFPSFEKEFQMQNSEVIS

  Query Match             3.5%;  Score 284.2;  DB 3;  Length 333321;
  Best Local Similarity   73.4%;  Pred. No. 3.7e-40;
  Matches  378;  Conservative    0;  Mismatches  133;  Indels    4;  Gaps    1;

Qy       7402 ACATCGGAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
              |||  |   | |||||||||||||||||||||||||||| |||||||||||||||||||||
Db      85220 ACAACAACAACAACAACAACAACAACAACAACAACAACCACAACAACAACAACAACAACA 85279

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85280 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 85339

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
              ||||||||||||||||||||||||| ||||| |||||||| || |||||||||||| ||
Db      85340 ACAACAACAACAACAACAACAACAACACCAACACCAACAACACCAGCAACAACAACACCA 85399

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
               |||||||||||||||||||||||||||||||||| || ||||||||||||||||||||||
Db      85400 GCAACAACAACAACAACAACAACAACAACAACACCAGCAACAACAACAACAACAACAACA 85459

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGA----CAACA 7697
              |||  ||||||||||||||||||||||||||||| || |||||||||||  |    |||||
Db      85460 ACACCAGCAACAACAACAACAACAACAACAACACCAGCAACAACAACACCAGCAGCAACA 85519

Qy       7698 ACAAGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAG 7757
```

```
                 |||| |  |  | ||    | |  || | ||    ||||   | | | ||  ||
Db       85520 ACAACAACAACACCAACAACACCAACACCAACAACACCAACAATACCAACAACAACAACA 85579

Qy        7758 CTTATGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGT 7817
                   |   || |  |   |   | | |   |||  || || |   | |  |  | ||
Db       85580 ACAACAACAACACCAACAATACCAACAACAACAACACCAACAATACCAACAACAACAACA 85639

Qy        7818 TCAAAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAA 7877
                 ||| |     || |  |     ||| ||||| || || || ||||| | ||| ||  |
Db       85640 ACAACAACAACATCAACAACAACACCAACACCAACACCAACAACACCAACAACACCAACA 85699

Qy        7878 ACCTCGAGTGTATGAGTGACGAGTACAACATGAAC 7912
                ||  | |    |  |   ||   |||||  |||
Db       85700 ACACCAACAACACCAACAACACCAACAACACCAAC 85734


RESULT 4
AC095632
LOCUS       AC095632               252215 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-8J8, WORKING DRAFT SEQUENCE.
ACCESSION   AC095632
VERSION     AC095632.6  GI:30467132
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 252215)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 252215)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 252215)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24940884.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GCWI
                Center clone name: CH230-8J8
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 234561 bases at least Q40
                Consensus quality: 236456 bases at least Q30
                Consensus quality: 237820 bases at least Q20
                Estimated insert size: 240560; sum-of-contigs estimation
                Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1   252215: contig of 252215 bp in length.
FEATURES             Location/Qualifiers
     source          1..252215
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-8J8"
     misc_feature    1..1377
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    6809..7694
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
```

```
                   end_sequence:BH346072"
     misc_feature  6867..7452
                   /note="clone_boundary
                   clone_end:T7
                   site:EcoRI
                   end_sequence:BH346072"
     misc_feature  244681..245527
                   /note="clone_boundary
                   clone_end:Sp6
                   site:EcoRI
                   end_sequence:BH346074"
     misc_feature  248849..249973
                   /note="wgs_end_extension
                   clone_end:T7"
     misc_feature  250825..252215
                   /note="wgs_end_extension
                   clone_end:T7"
ORIGIN

  Query Match             3.4%;  Score 276.6;  DB 2;  Length 252215;
  Best Local Similarity   74.9%;  Pred. No. 8.4e-39;
  Matches  359;  Conservative    0;  Mismatches  119;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
              ||| |    | || |||||||||||||||||||||||| ||||||||||||||| ||
Db     128578 ACAACAACAACAGCAACAACAACAACAACAACAACAACAGCAACAACAACAACAACAGCA 128637

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
              |||||||| ||||||||||||| |||||||||||||| ||||||||||||||||||| ||
Db     128638 ACAACAACAGCAACAACAACAACAGCAACAACAACAACAGCAACAACAACAACAACAGCA 128697

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
              ||||||||||| ||| |||||||||||||||| ||||||||||||| ||||||||||
Db     128698 ACAACAACAACAGCAATAACAACAACAACAACAGCAACAACAACAACAGCAACAACAACA 128757

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
              ||||||||||| ||||| |||||||||||||||| ||||||||||||| ||||||||
Db     128758 ACAACAACAACAGCAACAGCAACAACAACAACAACAGCAACAACAACAACAGCAACAACA 128817

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGA-CAACAACA 7700
              |||||||||||||| ||||||||||||| |||| ||||||||||| | ||||||||
Db     128818 ACAACAGCAACAACAGCAACAACAACAACAGCAACAGCAACAACAACAACAGCAACAACA 128877

Qy       7701 AGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTT 7760
              | |  |  | ||     |||  ||| | ||       |||||     | | ||  ||
Db     128878 ACAGCAACAACAACAGCAACAACAGCAACAGCAACAACAGCAACAACAACAACAACAGCA 128937

Qy       7761 ATGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCA 7820
              | | || |  |   || | | | || ||| |  | || |    |  |  |  ||  ||
Db     128938 ACAGCAACAACAACAACAGCAACAGCAACAACAACAACAACAGCAACAACAACAACAACA 128997

Qy       7821 AAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAAC 7879
                    |    |  | || |   || || || ||||| || || || | ||| ||  |||
Db     128998 GCAACAACAACAACAGCAGCAACAACAACAACAGCAACAGCAACAACAACAACAACAAC 129056

RESULT 5
AC096212
LOCUS       AC096212               229259 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-23N22, WORKING DRAFT SEQUENCE.
ACCESSION   AC096212
VERSION     AC096212.7  GI:30522384
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 229259)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229259)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 229259)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24818315.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
```

```
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
        -------------- Project Information
            Center project name: GERC
            Center clone name: CH230-23N22
        -------------- Summary Statistics
            Assembly program: Atlas 3.0;
            Consensus quality: 218284 bases at least Q40
            Consensus quality: 220932 bases at least Q30
            Consensus quality: 222479 bases at least Q20
            Estimated insert size: 225601; sum-of-contigs estimation
            Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
        --------------
        * NOTE: Estimated insert size may differ from sequence length
        *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 1 contigs. Gaps between the contigs
        * are represented as runs of N. The order of the pieces
        * is believed to be correct as given, however the sizes
        * of the gaps between them are based on estimates that have
        * provided by the submittor.
        * This sequence will be replaced
        * by the finished sequence as soon as it is available and
        * the accession number will be preserved.
        *          1  229259: contig of 229259 bp in length.
FEATURES             Location/Qualifiers
     source          1..229259
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-23N22"
     misc_feature    1..1447
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    complement(2726..3610)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH275579"
     misc_feature    complement(222929..223797)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH275580"
     misc_feature    224380..229259
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             3.4%;  Score 273.4;  DB 2;  Length 229259;
  Best Local Similarity   72.1%;  Pred. No. 3.1e-38;
  Matches  370;  Conservative    0;  Mismatches  141;  Indels    2;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db     117312 ACAACAACAGCAACAACAACAACAACAACAGCAACAACAACAACAGCAACAACAACAGCA 117371
Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db     117372 ACAACAACAACCACAACAACAACAGCAACAACAACAGCAACAACAACAACAAAGGCAACAACA 117431
Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db     117432 GCAACAACAGCAACAACAACAGCAACAACAACAACAGCAACAACAACAACAACAACAGCA 117491
Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db     117492 ACAACAACAACAACAACAGCAACAACAACAGCAACAGCAACAACAACAGCAACAACAACA 117551
Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db     117552 ACAGCAGCAACAACAACAACAACAGCAACAACAGCAACAACAACAACA--GCAGCAACAA 117609
Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db     117610 CAACAGCAACAGCAACAACAACAGCAACAACAACAACAGCAGCAACAACAACAACAGCAA 117669
Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db     117670 CAACAACAACAACAGCAACAACAGCAACAACAACAACAGCAGCAACAACAACAGCAACAA 117729
Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881
Db     117730 CAACAGCAACAACAACAGCAACAACAACAACAGCAGCAACAACAACAACAGCAACAACAA 117789
Qy       7882 CGAGTGTATGAGTGACGAGTACAACATGAACCG 7914
Db     117790 CGGCAATAACAACAACAGCAACAACAACAACAG 117822

RESULT 6
AC115666
LOCUS       AC115666              200412 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-261G15, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC115666
VERSION     AC115666.5  GI:25012557
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 200412)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
```

```
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200412)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 200412)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:23681737.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GTKP
                Center clone name: CH230-261G15
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 176292 bases at least Q40
                Consensus quality: 177912 bases at least Q30
                Consensus quality: 178895 bases at least Q20
                Estimated insert size: 176263; sum-of-contigs estimation
                Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1  149703: contig of 149703 bp in length
            *  149704  149803: gap of unknown length
            *  149804  194459: contig of 44656 bp in length
            *  194460  194559: gap of unknown length
            *  194560  199192: contig of 4633 bp in length
            *  199193  199292: gap of unknown length
            *  199293  200412: contig of 1120 bp in length.
FEATURES             Location/Qualifiers
     source          1..200412
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-261G15"
     misc_feature    1..1146
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    5088..6892
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(7420..8292)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ242807"
     misc_feature    15358..42115
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ242805"
     misc_feature    148445..149703
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    149804..150956
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    194560..195851
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN

  Query Match             3.4%;  Score 273.2;  DB 2;  Length 200412;
  Best Local Similarity   72.0%;  Pred. No. 3.4e-38;
  Matches  370;  Conservative    0;  Mismatches  143;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
              |||  |   |||||| |||||||||| ||||||| |||||||||| |||||||||
Db      60981 ACAGCAACAGCAACAGCAACAACAACAGCAACAACAGCAACAACAACAGCAACAACAACA 61040

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
               |||||||||| || ||||||||||||||| ||||||||||||||| ||||||||
Db      61041 GCAACAACAACAGCAGCAACAACAACAACAACAGCAACAACAACAACAACAGCAACAACA 61100

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
              ||||| |||||||||| ||||||||||||||| ||||||||||||||| ||||||||
Db      61101 ACAACAGCAACAACAACAGCAACAACAACAACAACAGCAACAACAACAACAGCAACAACA 61160

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
              ||| |||||||||| |||||||||| || ||||||||||||||| |||||||| ||
Db      61161 ACAGCAACAACAACAGCAACAACAACAGCAGCAACAACAACAACAACAGCAACAACAGCA 61220

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAA-GACAACAACA 7700
               |||| |||||||| ||||||| || ||||||||| |||| |||||| |||||||||
Db      61221 GCAACAACAACAACAGCAACAACAGCAGCAACAACAAGAACAGCAACAACAACAACAACA 61280

Qy       7701 AGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTT 7760
                |  |  | |||     |||  ||| | ||      ||||     ||| ||  ||
Db      61281 GCAACAACAACAGCAACAACAACAGCAACAGCAACAACAACAACAGCAACAACAGCAACA 61340

Qy       7761 ATGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCA 7820
                     || |  |  |  | ||  |||  |  | ||  |     |  |  |  ||  ||
Db      61341 GCAACAACAGCAACAACAACAACAGCAACAACAACAACAGCAACAACAACAACAACAGCA 61400

Qy       7821 AAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACC 7880
              |  |      |  |  ||      || || || ||||| || || || | |||||| |||
Db      61401 ACAACAACAGCAGCAGCAACAACAGCAACAACAGCAACAACAACAACAACAGCAACAACA 61460

Qy       7881 TCGAGTGTATGAGTGACGAGTACAACATGAACCG 7914
               |  |  |  ||  ||  |  ||||| ||| |
Db      61461 ACAACAGCAGCAGCAACAACAACAACAACAACAG 61494


RESULT 7
AC120325/c
LOCUS       AC120325              222913 bp    DNA     linear   HTG 15-NOV-2002
```

```
DEFINITION  Rattus norvegicus clone CH230-26G11, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC120325
VERSION     AC120325.6  GI:24635451
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 222913)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 222913)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 222913)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 6, 2002 this sequence version replaced gi:23265499.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
               Center project name: GWFX
               Center clone name: CH230-26G11
            -------------- Summary Statistics
               Assembly program: Phrap; version 0.990329
               Consensus quality: 216861 bases at least Q40
               Consensus quality: 218535 bases at least Q30
               Consensus quality: 219739 bases at least Q20
               Estimated insert size: 223729; sum-of-contigs estimation
               Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1  221799: contig of 221799 bp in length
            *   221800  221899: gap of unknown length
            *   221900  222913: contig of 1014 bp in length.
FEATURES             Location/Qualifiers
     source          1..222913
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-26G11"
     misc_feature    1..1281
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    2467..3303
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BH310471"
ORIGIN

  Query Match             3.4%;  Score 273.2;  DB 2;  Length 222913;
  Best Local Similarity   72.0%;  Pred. No. 3.4e-38;
  Matches  370;  Conservative    0;  Mismatches  143;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
              ||| | | |||||||||| |||||||| ||||||||||| ||||||||||||||||||||
Db     199900 ACAACAGCAGCAACAACAGCAACAACAGCAACAACAACAGCAACAACAACAACAACAACA 199841

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
               |||||||||||||| ||||||||||| ||||||||||||| |||||||||||| |||||
Db     199840 GCAACAACAACAACAGCAACAACAACAGCAACAACAACAACCACAACAACAACAGCAACA 199781

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
               |||||||||||||||||||| ||||||||||| ||||| ||||||||||| ||||||||
```

```
Db    199780 GCAACAACAACAACAACAACAGCAACAACAACAGCAACAGCAACAACAACAGCAACAACA 199721

Qy      7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
             |||||| || ||||||||||||||||| |||||||| |||||||||||||| || |||||
Db    199720 ACAACAGCAGCAACAACAACAACAACAGCAACAACAGCAACAACAACAACAGCAGCAACA 199661

Qy      7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAA-GACAACAACA 7700
             |||||||||||| ||||||||||| |||||||||||||| || ||||||  |||||| ||
Db    199660 ACAACAGCAACAGCAACAACAACAGCAACAACAACAACAGCAGCAACAACAACAACAGCA 199601

Qy      7701 AGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTT 7760
             | |  |  | ||     |||  ||| | ||     |||||    | | ||  ||
Db    199600 ACAACAACAACAACAGCAACAACAGCAACAACAACAACAGCAGCAACAACAACAGCAACA 199541

Qy      7761 ATGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCA 7820
             |    |  |  |   ||  | | |   |||  |  |  |     || |  |  ||   ||
Db    199540 ACAACAGCAACAACAACAGCAACAACAACAACAGCAGCAACAACAACAACAGCAACAACA 199481

Qy      7821 AAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACC 7880
             | |    |  |  |  | |   || || || ||||| || || || | ||| ||  |||
Db    199480 ACGGCAATAACAACAACAGCAACAACAACAACAGCAACAACAACAGCAACAACAGCAACA 199421

Qy      7881 TCGAGTGTATGAGTGACGAGTACAACATGAACCG 7914
              | |  | |  |   || |  ||| ||  ||| |
Db    199420 ACAACAGCAACAACAACAACAACAACAGCAACAG 199387
```

```
RESULT 8
AC117075
LOCUS       AC117075               254733 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 5201047-5455781 strain
            AX4, complete sequence.
ACCESSION   AC117075 AC117177 AC115615
VERSION     AC117075.2  GI:28828046
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 254733)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 254733)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 254733)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 254733)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology; Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 4, 2003 this sequence version replaced
            gi:20069690, gi:19570144, gi:20066242.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
```

```
FEATURES             Location/Qualifiers
     source          1..254733
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="5201047-5455781"
     CDS             complement(join(<124..619,719..1068,1167..1457))
                     /note="GeneID exon scores (in order of location ranges):
                     30.56, 24.41, 16.56 - GSCJ_ID dd_01823"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAM09355.2"
                     /db_xref="GI:28828135"
                     /translation="MGIEGLNLFMGKQFGCVEKVKEIPTDHVYIDLNNYLHKSVKRGD
                     NSKLNTEVNVFRILKSFIDGILRKVRVKHSVFFGIDGPGPRSKMILQRERRLKNGNID
                     KLKYFINRNKEQQQQQQQQQQQSPQLHDYDYLNENDLKSFDSSFSTLEFTPGTTFMGK
                     LKDFLIFYTKNKLYFAKKIFISASDRIGEGEFKIFEQILNSNYPINDSFTIVSSDSDI
                     LLFSLLSKYKNVYIYNKDSEEIIKIDKIRDKIYQQCYKKKQKKQKQQKKSGGEVEGEG
                     EGIATEEKEEGVQIEEEEEEEEEEEEDISKRRQAIVDFVMLTFLMGTDHLPKVSSYNI
                     GSAWSEYCKIKKPLYNEETGLINLDLLFRLIGKSSQPRLFYRNRF"
     CDS             complement(join(1796..3073,3249..3563))
                     /note="GeneID exon scores (in order of location ranges):
                     87.30, 9.45 - GSCJ_ID dd_01824"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO50731.1"
                     /db_xref="GI:28828047"
                     /translation="MKLGRDIPIYKLIIIKNESIENNLKEISFSDNFKCQICEGLLIS
                     SLIPNRMKALQCINGHCFCLTCWESILEIKSECPTCRIQIQSMNTLSNNLFIIKSISE
                     SIKIHCPNYLNFDNSNNFNGCKEIITIDEIDRHESKCEFRFIKCSINNQCNEIIRFNE
                     RDKHESQCDFIIQKCTHCDESVQMKQMQGHILFECLKVLITCQHCSLQFTRLKLQNHI
                     ENHCKEIKIECPFKSLKIIYNNNNNDNDNNDSDENNSNQSLSSSSLSSSIIPLSPCNE
                     LMKRSELSKHFIRNHQYHNELVSIVLKKQENKIKKLETIIQQNNVNQNFEISIIKHSN
                     SLTRDSFYKENKKKDKLIQDLMKRVLILEHQQQQNQIQNQIQNQIQKLSNKSIFTTTT
                     TTTSTSDKNNNNNNNNNNNNNNNNEDEEDDDNIKNNDNQGNLKSLREMFNTISNFKDD
                     ENEQQQIQQQQQLPFTSFGPLSVQQQNNPTKQFNQLSQPQTQPQSQSQSQSLFGEWSP
                     ITINQNQNTPSNPFSIFSGTSL"
     CDS             complement(join(4148..4536,4589..4883,4975..5261,
                     5357..5513,5597..5830,5946..6098))
                     /note="GeneID exon scores (in order of location ranges):
                     33.35, 9.09, 28.75, 10.40, 9.66, 11.24 - GSCJ_ID dd_01825"
                     /codon_start=1
                     /product="similar to Mus musculus (Mouse). 0 day neonate
                     eyeball cDNA, RIKEN full-length enriched library,
                     clone:E130103I15 product:polycystic kidney disease 2, full
                     insert sequence"
                     /protein_id="AAO50732.1"
                     /db_xref="GI:28828048"
                     /translation="MARSGYYIDIPVSDIKTGVESLIDNGFFDIQTRSVIISFSTLNL
                     NFQSRTSSGQHQDQLIPIISIRTYRIQLYMDALDKFRGFLEVSFFIFILYYFYYFVNE
                     ARIDFKLGRFKFYISNFKNIFDVINLTLFIASIALYLAFLGDNSRDLNINSQIQTTTD
                     YPLYLENLGQTALVLYQISAINILLMSFKTFRFLVVHRRLYILWIAISQSRLQLITFT
                     IMFCIMMLGFLFSGWLTFGTDIASFNGFVTSLGTLLQFIIGNPPDYEAMSYTNRALGP
                     IYYLLFTIFMFFILVNMFVAIISDSYQNISNTFEKKQKTKRYLLTRWERTMKSLLLLF
                     TPNSLDIYDLVLMFIDQKPGLIETDGIDPSRFKDEIMIIQPIQRARIKYLDQIQSAKN
                     TGVTFELKLMGQQVHKKDLDTWKKEAKIKEKKKFESIQQNINILNTKLDLLLQAMNVP
                     QQSIPQLINGSTGMPQSPIIINNNNNNGGSIPSSSSNGSINGSNHINLSGGTNTP"
     CDS             complement(join(6215..6478,6568..6864))
                     /note="GeneID exon scores (in order of location ranges):
                     21.78, 25.12 - GSCJ_ID dd_03502"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO50733.1"
                     /db_xref="GI:28828049"
                     /translation="MNTLKRTVFSQRLEAVKALENKTREEKATLLAAVETQRNARLVK
                     DLIIHLVFFVVFVLIVVLQLNPQSLFIINTSLKNQILFSQDSQYLEINNPSDFITYLN
                     STFCESISDMSLENNDAMNKIIGNVIRIRTARVKPDSCSNNGFNLTCYSNNYDKDTKD
                     RSPFGPNDMYTYTSNSHGSYVFGMFF"
     CDS             join(8148..8280,8374..10196)
                     /note="GeneID exon scores (in order of location ranges):
                     7.51, 117.32 - GSCJ_ID dd_00414"
```

```
                     /codon_start=1
                     /product="similar to Suppressor of mutant AC40 subunit of
                     RNA polymerase I and III (high serine); Srp40p
                     [Saccharomyces cerevisiae]"
                     /protein_id="AAO50734.1"
                     /db_xref="GI:28828050"
                     /translation="MNKTLSNKKASSKTTNNSNNNNNNTINDPVLSLPTYSLEIYEED
                     YFNELILEPLSFLKSGDVVITTSPTKKKTSSVSASPSPIKKESNSNSNSNSNSNSNSN
                     SNNNNNDSDSSSDNDLKFEGNQDDSEDEDQTNSNYYCPFCKKKFQTEQTWNTHINSSK
                     HLKSIKDQKKKSTTTGTSNTPTKNNNNNSPSNVKSPSSTKTKSSSSSSSSSSSSSSS
                     SSSSSSSSSSSSSSSSSSSSSSSSSSSSSTNSKDNEFLQDQIMQTQQAIKIKNSKP
                     NLACKNLFTIARSYCQKQYIRDSAKTLNLIIDILTISNNNNNNNTTPSITSTTTSSTS
                     NKSIVEYSLNNILNNATLLCKTHLYFGRLTRLFNQRLSEYHFLQCFNYLALINNDQLL
                     KYESSLVSSSPLNLYKIFIIQLDKFGKSKNKDTNTNTDTDTNTNTDTNKEIESIIQIN
                     KYINEIAGALCYCEVNFKSIILYILASCITFKFKLYKESLESNTRAFKLLSSLKQYHW
                     SIDLLLNNLLIKLKLNETIDIVDGDGDTEIQYQKLLLKEKEEKEEKEYQNIIIILIKY
                     CILIDDKIRLSAILGDYNFQKIISIYFKNNDYVDCLIKSSNQFLIENYKEIDQIKDHY
                     EHLIYSNKDFIFIFNQFDLKKKILLQK"
     CDS             complement(join(10448. .10915,11145. .11294))
                     /note="GeneID exon scores (in order of location ranges):
                     45.58, 25.53 - GSCJ_ID dd_00412"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Ras-related protein Rap-1"
                     /protein_id="AAO50735.1"
                     /db_xref="GI:28828051"
                     /translation="MGKGNGKSKDSKGSEKSSTKTKFFVAVMGAGSVGKSALTVQFTQ
                     GVFIDKYDPTVEDTYTKTFELDGEQVCIEVLDTAGSEVLVAMRELYMKNAEGFILVYS
                     ILVKSTFLEIKDIVEQLFRVKEEEEVPIVLVGNKIDLDAHREVSTNEGKQLANSYPNC
                     DFWEASSKDRINVDNVFFSIVKRIKEKYKKEGAPVKEKKEKCIIM"
     CDS             12794. .13423
                     /note="GeneID exon scores (in order of location ranges):
                     35.19 - GSCJ_ID dd_00410"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO50736.1"
                     /db_xref="GI:28828052"
                     /translation="MASIGTGKFKNEPTFGKNIKEILILLERSIAESEKQWVDLVSEN
                     KNLVDVNAFRINKLFDNKIGLDVKDNKSFELLKKTVHEFKTERPVIDLVCKSISCLFY
                     LSDLQLSSIITNLPNEIIAQLKDSTVPFAYRLENSSLKNVQIQKRHYLSSNRPIELVC
                     CILYLPANIDIICKVVPKSKFLNETSISGCPFDLEMDSSFYGIKKKLSL"
     CDS             complement(join(13494. .13813,13915. .14097,14208. .14241))
                     /note="GeneID exon scores (in order of location ranges):
                     22.37, 12.24, 8.90 - GSCJ_ID dd_00408"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO50737.1"
                     /db_xref="GI:28828053"
                     /translation="MSIYEEIMRAQVKQGNMGGRSGPLGYLLRENKQKEKKAVKVNTP
                     KPPKPVVTQSPYTINLIEGTVTNATDNNGFSDPYLVIIAIKPDNSFSKILFKSNVCKK
                     TLTPTWNQVENFFVKDNVTDIIVELWDHDVLSKNDFIGRSIISMNTVRSKGYDGTLEF
                     FNGSEVVDGTVKLTIKRK"
     CDS             complement(join(14655. .14686,14848. .14916,14992. .15162,
                     15268. .15310))
                     /note="GeneID exon scores (in order of location ranges):
                     1.29, 3.37, 1.88, 4.01 - GSCJ_ID dd_00406"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO50738.1"
                     /db_xref="GI:28828054"
                     /translation="MSEICSTSCWQYAIDNPNCFPKIGFTKVCFMSGNYTCASSSGFT
                     ILSDELSFINYGLYHPDEIIGKFIKIKGNEYKVTKATNSIINLSNSKNGRVTCLPKRR
                     VV"
     CDS             complement(join(15876. .15900,16018. .17390))
                     /note="GeneID exon scores (in order of location ranges):
                     -0.56, 167.37 - GSCJ_ID dd_00405"
                     /codon_start=1
                     /product="similar to Physarum polycephalum (Slime mold).
                     Glutathione reductase"
                     /protein_id="AAM09354.2"
                     /db_xref="GI:28828055"
                     /translation="MSSTNHFTYLVLGAGSGGIASARRAAKHLNAKGNGDRIGIVEVT
                     RPGGTCVNVGCVPKKVMWNTSFIKEMINAAPSYGFDFGGQQVKFNWPTIKKARDEYIK
                     RLNGIYDSNLAKDNIVRINGYGRFSGPKEIQVNGANGEKYTADHILIAAGGRPTVPDV
                     PGKELGITSDGFFELEDLPKSTLVVGAGYIAVELAGVLHSLGSETTMVIRQKQFLRTF
                     DEMLHTTLLKQMTDDGVKFVTEASIKSLERDVDGKRIIATTNAGVKLPPVECVIWAIG
                     RVPNTDDLGIDKAGIQLTEQSGFIKVDEFQNTNVPGVHAVGDICGNFLLTPVAIAAGR
                     RLSERLFNGKSDLKFEYENVATVVFSHPPIGTVGLTEQEAITKYGTENIKCYNTSFIN
                     MFYSVQVHKVRTSMKLVCLGKEEKVIGLHIIGDGCDEIIQGFAVAVKMGCTKWDLDNT
                     CAIHPTSAEELVTMV"
     CDS             complement(join(18106. .18501,18638. .18994))
                     /note="GeneID exon scores (in order of location ranges):
                     56.32, 49.02 - GSCJ_ID dd_00404"

  Query Match             3.4%;  Score 271.4;  DB 3;  Length 254733;
  Best Local Similarity   68.6%;  Pred. No. 6.9e-38;
  Matches  389;  Conservative    0;  Mismatches  176;  Indels    2;  Gaps    1;

Qy       7399 ATCACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7458
Db     161652 ACCACATCAACATCAACAACAACAACATCAACAACCACATCAACAACCACATCATCAACA 161711

Qy       7459 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7518
Db     161712 ACAACACCACCAAAACAATCATCAACAACATCAACAACATCAACAACACCAACATCATCA 161771

Qy       7519 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7578
Db     161772 ACAACAACAACTACAACAACAACAACAACAACAACAACAACAACAACACCACCAACATCA 161831

Qy       7579 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7638
Db     161832 ACAACATCATCATCATCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 161891

Qy       7639 ACAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAA 7698
Db     161892 ACAACAACAACAACAACAACAACAACAACAACACCAACAACAACAACA--ACAACAA 161949

Qy       7699 CAAGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGC 7758
Db     161950 CAACAACAACACCAACACCAACAACAACACCAACACCAACAACACCAACACCAACAACAC 162009

Qy       7759 TTATGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTT 7818
Db     162010 CAACACCAACACCAACAACACCAACACCAACACCAACACCAACACCAACACCAACACCAA 162069

Qy       7819 CAAAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAA 7878
Db     162070 CACCAACACCAACACCAACACCAACACCAACACCAACACCAACACCAACACCAACATCAA 162129

Qy       7879 CCTCGAGTGTATGAGTGACGAGTACAACATGAACCGTCGACGATCCCTATAGTGAGTCGT 7938
Db     162130 CATCAACAACATCAACAACATCAACAACATCAACAACAACATCACCATCAACAAAATCAA 162189

Qy       7939 ATTATGCGGCCGCGAATTCGATATCAA 7965
Db     162190 CACCACCATCAACAAAATCAACATCAA 162216


RESULT 9
AC141123/c
LOCUS       AC141123               118689 bp    DNA     linear   HTG 27-MAR-2003
DEFINITION  Rattus norvegicus clone CH230-457B19, *** SEQUENCING IN PROGRESS
            ***, 57 unordered pieces.
ACCESSION   AC141123
VERSION     AC141123.1  GI:28893600
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 118689)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
```

```
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 118689)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 118689)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KDZH
                Center clone name: CH230-457B19
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 101909 bases at least Q40
                Consensus quality: 108962 bases at least Q30
                Consensus quality: 114915 bases at least Q20
                Estimated insert size: 98018; sum-of-contigs estimation
                Quality coverage: 1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 57 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1487: contig of 1487 bp in length
            *     1488     1587: gap of unknown length
            *     1588     2907: contig of 1320 bp in length
            *     2908     3007: gap of unknown length
            *     3008     4287: contig of 1280 bp in length
            *     4288     4387: gap of unknown length
            *     4388     5390: contig of 1003 bp in length
            *     5391     5490: gap of unknown length
            *     5491     6709: contig of 1219 bp in length
            *     6710     6809: gap of unknown length
            *     6810     7985: contig of 1176 bp in length
            *     7986     8085: gap of unknown length
            *     8086     9672: contig of 1587 bp in length
            *     9673     9772: gap of unknown length
            *     9773    11010: contig of 1238 bp in length
            *    11011    11110: gap of unknown length
            *    11111    12167: contig of 1057 bp in length
            *    12168    12267: gap of unknown length
            *    12268    13860: contig of 1593 bp in length
            *    13861    13960: gap of unknown length
            *    13961    15082: contig of 1122 bp in length
            *    15083    15182: gap of unknown length
            *    15183    16254: contig of 1072 bp in length
            *    16255    16354: gap of unknown length
            *    16355    17408: contig of 1054 bp in length
            *    17409    17508: gap of unknown length
            *    17509    18905: contig of 1397 bp in length
            *    18906    19005: gap of unknown length
            *    19006    20477: contig of 1472 bp in length
            *    20478    20577: gap of unknown length
            *    20578    21753: contig of 1176 bp in length
            *    21754    21853: gap of unknown length
            *    21854    23351: contig of 1498 bp in length
            *    23352    23451: gap of unknown length
            *    23452    24738: contig of 1287 bp in length
            *    24739    24838: gap of unknown length
            *    24839    25912: contig of 1074 bp in length
            *    25913    26012: gap of unknown length
            *    26013    27129: contig of 1117 bp in length
            *    27130    27229: gap of unknown length
            *    27230    28758: contig of 1529 bp in length
            *    28759    28858: gap of unknown length
            *    28859    30576: contig of 1718 bp in length
            *    30577    30676: gap of unknown length
            *    30677    31803: contig of 1127 bp in length
            *    31804    31903: gap of unknown length
            *    31904    32903: contig of 1000 bp in length
            *    32904    33003: gap of unknown length
            *    33004    34218: contig of 1215 bp in length
            *    34219    34318: gap of unknown length
            *    34319    35622: contig of 1304 bp in length
            *    35623    35722: gap of unknown length
            *    35723    37153: contig of 1431 bp in length
            *    37154    37253: gap of unknown length
            *    37254    38992: contig of 1739 bp in length
            *    38993    39092: gap of unknown length
            *    39093    40732: contig of 1640 bp in length
            *    40733    40832: gap of unknown length
            *    40833    42126: contig of 1294 bp in length
```

```
   *     42127     42226: gap of unknown length
   *     42227     43552: contig of 1326 bp in length
   *     43553     43652: gap of unknown length
   *     43653     44711: contig of 1059 bp in length
   *     44712     44811: gap of unknown length
   *     44812     47103: contig of 2292 bp in length
   *     47104     47203: gap of unknown length
   *     47204     48836: contig of 1633 bp in length
   *     48837     48936: gap of unknown length
   *     48937     50830: contig of 1894 bp in length
   *     50831     50930: gap of unknown length
   *     50931     52821: contig of 1891 bp in length
   *     52822     52921: gap of unknown length
   *     52922     54182: contig of 1261 bp in length
   *     54183     54282: gap of unknown length
   *     54283     56354: contig of 2072 bp in length
   *     56355     56454: gap of unknown length
   *     56455     57714: contig of 1260 bp in length
   *     57715     57814: gap of unknown length
   *     57815     60151: contig of 2337 bp in length
   *     60152     60251: gap of unknown length
   *     60252     63938: contig of 3687 bp in length
   *     63939     64038: gap of unknown length
   *     64039     66503: contig of 2465 bp in length
   *     66504     66603: gap of unknown length
   *     66604     69921: contig of 3318 bp in length
   *     69922     70021: gap of unknown length
   *     70022     72039: contig of 2018 bp in length
   *     72040     72139: gap of unknown length
   *     72140     74920: contig of 2781 bp in length
   *     74921     75020: gap of unknown length
   *     75021     77604: contig of 2584 bp in length
   *     77605     77704: gap of unknown length
   *     77705     80703: contig of 2999 bp in length
   *     80704     80803: gap of unknown length
   *     80804     84177: contig of 3374 bp in length
   *     84178     84277: gap of unknown length
   *     84278     87268: contig of 2991 bp in length
   *     87269     87368: gap of unknown length

 Query Match             3.4%;  Score 271;  DB 2;  Length 118689;
 Best Local Similarity   70.1%;  Pred. No. 8.5e-38;
 Matches  378;  Conservative    0;  Mismatches  160;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db     109211 ACAACAGCAACAACAACAGCAACAACAACAGCAACAACAACAGCAACAGCAACAACAACA 109152

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db     109151 GCAACAACAACAGCAACAACAACAACAGCAACAACAACAGCAGCAACAGCAACAACAACA 109092

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db     109091 GCAACAACAACAGCAGCAACAACAACAACAACAGCAACAACAACAGCAGCAACAACAACA 109032

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db     109031 ACAGCAACAGCAACAACAACAGCAACAACAACAACAGCAACAACAACAACAACAGCAACA 108972

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACA-AGACAACAACA 7700
Db     108971 ACAACAGCAGCAACAACAACAGCAACAACAACAGCAACAACAACAACAGCAACAACAACA 108912

Qy       7701 AGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTT 7760
Db     108911 ACAACAACAGCAACAACAACAACAGCAACAGCAACAGCAACAACAACAACAGCAGCAACA 108852

Qy       7761 ATGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCA 7820
Db     108851 ACAGCAACAACAACAGCAACAACAACAACAGCAACAGCAACAACAACAGCAGCAACAACA 108792

Qy       7821 AAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACC 7880
Db     108791 ACAACAGCAACAACAACAACAGCAACAACAACAGCAACAACAACAGCAACAACAACAACA 108732

Qy       7881 TCGAGTGTATGAGTGACGAGTACAACATGAACCGTCGACGATCCCTATAGTGAGTCGTA 7939
Db     108731 ACAACAGCAGCAACAACAACAGCAACAACAACAATCTACTATCATTTAATTTAGTAATA 108673


RESULT 10
AC096792/c
LOCUS       AC096792               205454 bp    DNA     linear   ROD 31-AUG-2003
DEFINITION  Rattus norvegicus 7 BAC CH230-8015 (Children's Hospital Oakland
            Research Institute) complete sequence.
ACCESSION   AC096792
VERSION     AC096792.12  GI:34368615
KEYWORDS    HTG.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 205454)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 205454)
```

```
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 205454)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 205454)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-AUG-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Aug 31, 2003 this sequence version replaced gi:30467011.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.
FEATURES             Location/Qualifiers
     source          1..205454
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /chromosome="7"
                     /clone="CH230-8O15"
     repeat_region   142..182
                     /rpt_family="(GTTTG)n"
     repeat_region   complement(2365..3271)
                     /rpt_family="L1_Rat3"
     repeat_region   3271..3712
                     /rpt_family="L1_Rat3"
     repeat_region   3713..4323
                     /rpt_family="L1_Rn"
     STS             3862..4084
                     /standard_name="RH138728"
     repeat_region   4324..4619
                     /rpt_family="L1_Rat3"
     repeat_region   4658..4692
                     /rpt_family="(CCAA)n"
     repeat_region   4718..4987
                     /rpt_family="B4A"
     repeat_region   5233..5310
                     /rpt_family="(TCTA)n"
     repeat_region   complement(7758..7837)
                     /rpt_family="Lx3C"
     repeat_region   7838..7888
                     /rpt_family="(CA)n"
     repeat_region   complement(7889..8124)
                     /rpt_family="Lx3C"
     repeat_region   complement(9102..9218)
                     /rpt_family="BC1_Rn"
     repeat_region   10199..10275
                     /rpt_family="(TG)n"
     repeat_region   complement(10276..10430)
                     /rpt_family="B3A"
     repeat_region   11359..11415
                     /rpt_family="(GA)n"
     repeat_region   11853..11888
                     /rpt_family="(GGGA)n"
     repeat_region   12009..12074
                     /rpt_family="(TCCC)n"
     repeat_region   complement(12162..12838)
                     /rpt_family="Lx7"
     repeat_region   14382..14413
                     /rpt_family="(TAAA)n"
     repeat_region   16045..16348
                     /rpt_family="Lx7"
     repeat_region   complement(17248..17457)
                     /rpt_family="L1MA9"
     STS             17436..17659
                     /standard_name="AU047209"
     repeat_region   17458..17513
                     /rpt_family="(TG)n"
     repeat_region   complement(17514..17628)
                     /rpt_family="L1MA9"
     repeat_region   complement(17840..17974)
                     /rpt_family="BC1_Rn"
     repeat_region   19067..19195
                     /rpt_family="(TTAA)n"
     repeat_region   20409..20621
                     /rpt_family="MER99"
     repeat_region   20860..20937
                     /rpt_family="(TCTA)n"
     repeat_region   complement(20959..21267)
                     /rpt_family="Lx7"
     repeat_region   complement(21838..22819)
                     /rpt_family="Lx8"
     repeat_region   complement(22952..23799)
                     /rpt_family="RMER13B"
     repeat_region   25048..25228
                     /rpt_family="B4A"
     repeat_region   25357..25545
                     /rpt_family="B4A"
     repeat_region   25565..25597
                     /rpt_family="(CA)n"
     repeat_region   26421..26598
                     /rpt_family="RLTR40"
     repeat_region   26676..26894
                     /rpt_family="B3"
     repeat_region   29766..29789
                     /rpt_family="(TA)n"
     repeat_region   29902..30034
                     /rpt_family="B1F"
     repeat_region   31226..31253
                     /rpt_family="(ATTG)n"
     repeat_region   complement(32491..32622)
                     /rpt_family="ID_Rn"
     repeat_region   32656..32802
                     /rpt_family="(TCTA)n"
     repeat_region   32898..32992
                     /rpt_family="CT-rich"
     repeat_region   33041..33103
                     /rpt_family="(TA)n"
     repeat_region   complement(33112..33168)
                     /rpt_family="B1F"
     repeat_region   34236..34267
                     /rpt_family="AT_rich"
     repeat_region   complement(35663..35902)
                     /rpt_family="L1_Mur3"
     repeat_region   complement(37158..37277)
                     /rpt_family="RMER17C"
     repeat_region   37278..37457
                     /rpt_family="(GGA)n"
     repeat_region   complement(37458..37471)
                     /rpt_family="RMER17C"
     repeat_region   complement(37905..38038)
                     /rpt_family="B1F"
     repeat_region   38063..38221
                     /rpt_family="URR1B"
     repeat_region   38408..38491
                     /rpt_family="GA-rich"
     repeat_region   40283..40380

  Query Match             3.4%;  Score 271;  DB 10;  Length 205454;
  Best Local Similarity   70.1%;  Pred. No. 8.3e-38;
  Matches  378;  Conservative    0;  Mismatches  160;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
              ||| | | | |||||||| ||||||||||| ||||||||||| ||||| |||||||||||
Db     189671 ACAACAGCAACAACAACAGCAACAACAACAGCAACAACAACAGCAACAGCAACAACAACA 189612
```

```
Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db     189611 GCAACAACAACAGCAACAACAACAACAGCAACAACAACAGCAGCAACAGCAACAACAACA 189552
Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db     189551 GCAACAACAACAGCAGCAACAACAACAACAACAGCAACAACAACAGCAGCAACAACAACA 189492
Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db     189491 ACAGCAACAGCAACAACAACAGCAACAACAACAACAGCAACAACAACAACAACAGCAACA 189432
Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACA-AGACAACAACA 7700
Db     189431 ACAACAGCAGCAACAACAACAGCAACAACAACAGCAACAACAACAACAGCAACAACAACA 189372
Qy       7701 AGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTT 7760
Db     189371 ACAACAACAGCAACAACAACAACAGCAACAGCAACAGCAACAACAACAACAGCAGCAACA 189312
Qy       7761 ATGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCA 7820
Db     189311 ACAGCAACAACAACAGCAACAACAACAACAGCAACAGCAACAACAACAGCAGCAACAACA 189252
Qy       7821 AAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACC 7880
Db     189251 ACAACAGCAACAACAACAACAGCAACAACAACAGCAACAACAACAGCAACAACAACAACA 189192
Qy       7881 TCGAGTGTATGAGTGACGAGTACAACATGAACCGTCGACGATCCCTATAGTGAGTCGTA 7939
Db     189191 ACAACAGCAGCAACAACAACAGCAACAACAACAATCTACTATCATTTAATTTAGTAATA 189133
```

```
RESULT 11
AC110118
LOCUS       AC110118             198573 bp    DNA     linear   HTG 30-OCT-2003
DEFINITION  Rattus norvegicus clone CH230-310N6, WORKING DRAFT SEQUENCE, 13
            unordered pieces.
ACCESSION   AC110118
VERSION     AC110118.7  GI:38044116
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 198573)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198573)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 198573)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 30, 2003 this sequence version replaced gi:32401502.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GQJY
                Center clone name: CH230-310N6
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 190090 bases at least Q40
                Consensus quality: 191834 bases at least Q30
                Consensus quality: 192877 bases at least Q20
                Estimated insert size: 205775; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *    (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
```

```
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *        1       2085: contig of 2085 bp in length
           *     2086       2577: gap of 492 bp
           *     2578      68887: contig of 66310 bp in length
           *    68888      68937: gap of 50 bp
           *    68938      69018: contig of 81 bp in length
           *    69019      69073: gap of 55 bp
           *    69074     145466: contig of 76393 bp in length
           *   145467     145537: gap of 71 bp
           *   145538     145540: contig of 3 bp in length
           *   145541     145590: gap of 50 bp
           *   145591     181558: contig of 35968 bp in length
           *   181559     183949: gap of 2391 bp
           *   183950     187285: contig of 3336 bp in length
           *   187286     187385: gap of unknown length
           *   187386     188812: contig of 1427 bp in length
           *   188813     189535: gap of 723 bp
           *   189536     191742: contig of 2207 bp in length
           *   191743     191842: gap of unknown length
           *   191843     193197: contig of 1355 bp in length
           *   193198     193297: gap of unknown length
           *   193298     194379: contig of 1082 bp in length
           *   194380     194479: gap of unknown length
           *   194480     196272: contig of 1793 bp in length
           *   196273     196372: gap of unknown length
           *   196373     198573: contig of 2201 bp in length.
FEATURES             Location/Qualifiers
     source          1. .198573
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-310N6"
     misc_feature    1. .2085
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(28893. .28925)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:MboI
                     end_sequence:BZ148934"
     misc_feature    complement(179991. .180840)
                     /note="clone_boundary
                     clone_end:T7
                     site:MboI
                     end_sequence:BZ148932"
     misc_feature    183950. .187285
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    187386. .188812
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN

  Query Match             3.3%;  Score 269.2;  DB 2;  Length 198573;
  Best Local Similarity   71.8%;  Pred. No. 1.7e-37;
  Matches  366;  Conservative    0;  Mismatches  143;  Indels    1;  Gaps    1;

Qy       7399 ATCACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7458
              || ||| | |  ||||||||||||| |||||||||||| |||||||||||||| |||||
Db      19211 ATAACAACAGCAACAACAACAACAACAGCAACAACAACAGCAACAACAACAACAGCAACA 19270

Qy       7459 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7518
               | |||||||| ||||||||||||| ||||||| |||||||||| |||||||||||||
Db      19271 AAAACAACAAAAACAACAACAACAATAACAACAGCAACAACAACAGCAACAACAACAACA 19330

Qy       7519 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7578
              ||| |||||| ||||||||||||||||| |||||| |||||||||||||||||||||
Db      19331 ACAGCAACAAAAACAACAACAACAACAACAGCAACAAAAACAACAACAACAACAACAACA 19390

Qy       7579 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7638
                |||||||  |||||||||||||||| |||||||  ||||||||||| ||||||| ||
Db      19391 GCAACAACAGTAACAACAACAACAACAAAAACAACAGTAACAACAACAATAACAACAGCA 19450

Qy       7639 ACAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGA-CAACA 7697
              ||||||||||||||||||||||| |||||||| ||||||||||||||| | |||||
Db      19451 ACAACAACAGCAACAACAACAACAGCAACAACAGTAACAACAACAACAACAACAGCAACA 19510

Qy       7698 ACAAGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAG 7757
              |||  |  |  | ||      |||  ||   | ||   | ||||        |  |  || |
Db      19511 ACAGTAACAACAACAATAACAACAGCAACAACAACAGCAACAACAACAACAGCAACAAAA 19570

Qy       7758 CTTATGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGT 7817
                 |  | || |   |    ||  | | |   ||  |  | ||  |    |    |  ||
Db      19571 ACAACAGCAACAACAATAACAACAACAACAGCAACAACAACAACAACAGCAACAACAACA 19630

Qy       7818 TCAAAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAA 7877
               |||| |      |  |   | ||  || || ||  | || || || || |  ||| ||  |
Db      19631 ACAAAAACAACAGCAACAACAGTAACAACAACAATAACAACAGCAACAACAACAACAACA 19690

Qy       7878 ACCTCGAGTGTATGAGTGACGAGTACAACA 7907
               | | |    |  ||  || |  |||||
Db      19691 GCAACAAAAACAACAGCAACAACAACAACA 19720


RESULT 12
AC099432
LOCUS       AC099432              255678 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-103G20, *** SEQUENCING IN PROGRESS
            ***, 9 unordered pieces.
ACCESSION   AC099432
VERSION     AC099432.5  GI:24942644
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 255678)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
```

```
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 255678)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 255678)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:23268717.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHWE
                Center clone name: CH230-103G20
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 229162 bases at least Q40
                Consensus quality: 232609 bases at least Q30
                Consensus quality: 234944 bases at least Q20
                Estimated insert size: 236379; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     24634: contig of 24634 bp in length
            *    24635     24734: gap of unknown length
            *    24735    166364: contig of 141630 bp in length
            *   166365    166464: gap of unknown length
            *   166465    172815: contig of 6351 bp in length
            *   172816    172915: gap of unknown length
            *   172916    191173: contig of 18258 bp in length
```

```
            *   191174    191273: gap of unknown length
            *   191274    249873: contig of 58600 bp in length
            *   249874    249973: gap of unknown length
            *   249974    251258: contig of 1285 bp in length
            *   251259    251358: gap of unknown length
            *   251359    253012: contig of 1654 bp in length
            *   253013    253112: gap of unknown length
            *   253113    254476: contig of 1364 bp in length
            *   254477    254576: gap of unknown length
            *   254577    255678: contig of 1102 bp in length.
FEATURES             Location/Qualifiers
     source          1..255678
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-103G20"
     misc_feature    161290..162357
                     /note="wgs_contig"
     misc_feature    166465..169028
                     /note="wgs_contig"
     misc_feature    169079..172815
                     /note="wgs_contig"
     misc_feature    172916..173975
                     /note="wgs_contig"
     misc_feature    191274..192552
                     /note="wgs_contig"
ORIGIN

  Query Match              3.3%;  Score 268.6;  DB 2;  Length 255678;
  Best Local Similarity   71.5%;  Pred. No. 2.2e-37;
  Matches  367;  Conservative    0;  Mismatches  144;  Indels    2;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461

Db     186703 ACAGCAACAGCAACAACAGCAACAGCAACAACAACAGCAACAACAACAACAGCAACAGCA 186762

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521

Db     186763 ACAACAACAACAGCAACAGCAACAACAACAACAACAACAGCAACAGCAACAACAACAACA 186822

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581

Db     186823 GCAACAACAACAGCAACAGCAACAACAACAACAGCAACAACAACAACAGCAACAGCAACA 186882

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641

Db     186883 GCAACAACAGCAACAGCAACAGCAACAACAGCAACAACAACAACAGCAACAACAACAACA 186942

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701

Db     186943 GCAACAGCAACAGCAACAGCAACAACAACAGCAACAGCAACAACAACA--ACAGCAACAA 187000

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761

Db     187001 CAACAACAGCAACAACAACAACAGCAACAGCAACAACAGCAACAGCAACAACAACAGCAA 187060

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821

Db     187061 CAGCAACAGCAACAACAACAACAGCAACAACAGCAACAGCAACAACAGCAACAACAACAA 187120

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881

Db     187121 CAGCAACAACAACAACAGCAACAACAACAGCAACAGCAACAACAACAACAGCAACAACAG 187180

Qy       7882 CGAGTGTATGAGTGACGAGTACAACATGAACCG 7914

Db     187181 CAACAGCAACAACAACAGCAACAACAGCAACAG 187213


RESULT 13
AC114079
LOCUS       AC114079               279242 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-8H10, *** SEQUENCING IN PROGRESS ***,
```

```
            4 unordered pieces.
ACCESSION   AC114079
VERSION     AC114079.4  GI:30580483
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 279242)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 279242)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 279242)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23268757.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GLHC
                Center clone name: CH230-8H10
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 218357 bases at least Q40
                Consensus quality: 224314 bases at least Q30
                Consensus quality: 228338 bases at least Q20
                Estimated insert size: 230439; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   261372: contig of 261372 bp in length
            *   261373   261472: gap of unknown length
            *   261473   276851: contig of 15379 bp in length
            *   276852   276951: gap of unknown length
            *   276952   278133: contig of 1182 bp in length
            *   278134   278233: gap of unknown length
            *   278234   279242: contig of 1009 bp in length.
FEATURES             Location/Qualifiers
     source          1..279242
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-8H10"
     misc_feature    1..3187
                     /note="wgs_contig"
     misc_feature    194262..196958
                     /note="wgs_contig"
     misc_feature    214939..216212
                     /note="wgs_contig"
     misc_feature    261473..262788
                     /note="wgs_contig"
ORIGIN

  Query Match             3.3%;  Score 268.4;  DB 2;  Length 279242;
  Best Local Similarity   71.3%;  Pred. No. 2.3e-37;
  Matches  368;  Conservative    0;  Mismatches  146;  Indels    2;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
              |||  |   | |||||  ||||||||||||| |||||||||||  |||||||||||||||
Db     124032 ACAACAACAACAACAGCAACAACAACAACAGCAACAACAACAGCAACAACAACAACAACA 124091

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
               ||||||||||||| |||||||||||  |||||||||||  |||||||||||  || |||||
Db     124092 GCAACAACAACAACAGCAACAACAACAGCAACAACAACAGCAACAACAACAGCAGCAACA 124151
```

```
Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db     124152 ACAACAACAACAGCAACAACAGCAGCAGCAGCAACAACAACAACAACAGCAACAACAGCA 124211

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db     124212 ACAACAACAGCAACAACAACAACAACAGCAACAACAACAGCAACAACAACAACAACAGCA 124271

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db     124272 ACAACAGCAACAACAACAGCAACAACAACAACAGCAACAACAGCAACA--ACAACAGCAA 124329

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db     124330 CAACAACAACAGCAACAACAACAACAGCAACAACAACAGCAGCAGCAACAACAGCAACAA 124389

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db     124390 CAACAACAGCAACAACAACAACAGCAACAACAACAGCAACAACAACAACAGCAACAACAA 124449

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881
Db     124450 CAGCAACAACAACAACAGCAACAACAACAACAACAGCAACAGCAACAACAGCAACAACAA 124509

Qy       7882 CGAGTGTATGAGTGACGAGTACAACATGAACCGTCG 7917
Db     124510 CAACAGCAACAACAACAGCAACAACAACAACAATGG 124545
```

```
RESULT 14
AC128290
LOCUS       AC128290               346208 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-303N2, WORKING DRAFT SEQUENCE.
ACCESSION   AC128290
VERSION     AC128290.3  GI:24635394
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 346208)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 346208)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 346208)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 6, 2002 this sequence version replaced gi:23907971.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GZHE
                Center clone name: CH230-303N2
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 199259 bases at least Q40
                Consensus quality: 201097 bases at least Q30
                Consensus quality: 201885 bases at least Q20
                Estimated insert size: 205268; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *    (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
```

```
         *            1   346208: contig of 346208 bp in length.
FEATURES              Location/Qualifiers
     source           1. .346208
                      /organism="Rattus norvegicus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10116"
                      /clone="CH230-303N2"
     misc_feature     1. .1207
                      /note="wgs_end_extension
                      clone_end:Sp6"
     misc_feature     137411. .139197
                      /note="wgs_end_extension
                      clone_end:Sp6"
     misc_feature     144420. .145263
                      /note="clone_boundary
                      clone_end:Sp6
                      site:
                      end_sequence:BZ202310"
     misc_feature     complement(342409. .343270)
                      /note="clone_boundary
                      clone_end:T7
                      site:
                      end_sequence:BZ202308"
ORIGIN

  Query Match             3.3%;  Score 268.4;  DB 2;  Length 346208;
  Best Local Similarity   71.4%;  Pred. No. 2.3e-37;
  Matches  367;  Conservative    0;  Mismatches  146;  Indels    1;  Gaps    1;

Qy        7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
               ||| |   ||||||| ||||||||||| |||||||| ||||||||||| |||||||||||
Db      288972 ACAGCAACAGCAACAGCAACAACAACAGCAACAACAGCAACAACAACAGCAACAACAACA 289031

Qy        7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
                ||||||||||| || ||||||||||||||||| ||||||||||||||||| ||||||||
Db      289032 GCAACAACAACAGCAGCAACAACAACAACAACAGCAACAACAACAACAACAGCAACAACA 289091

Qy        7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
               |||||| ||||||||||| ||||||||||||||||| |||||||||||||| ||||||||
Db      289092 ACAACAGCAACAACAACAGCAACAACAACAACAACAGCAACAACAACAACAGCAACAACA 289151

Qy        7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
               ||| ||||||||||| ||||||||||| || ||||||||||||||||| |||||||| ||
Db      289152 ACAGCAACAACAACAGCAACAACAACAGCAGCAACAACAACAACAACAGCAACAACAGCA 289211

Qy        7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAA-GACAACAACA 7700
                ||||| |||||||| |||||||| || ||||||||| |||| ||||||  |||||||||
Db      289212 GCAACAACAACAACAGCAACAACAGCAGCAACAACAAGAACAGCAACAACAACAACAACA 289271

Qy        7701 AGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTT 7760
                 |  |  | |||    |||  ||| | ||     ||||     ||| ||  ||
Db      289272 GCAACAACAACAGTAACAACAACAGCAACAGCAACAACAACAACAGCAACAACAGCAACA 289331

Qy        7761 ATGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCA 7820
                |   ||    |   ||  | |  | |||   |  || |     |  |  |  ||   ||
Db      289332 GTAACAACTGCAACAACTACTACTGCTACTATAACAACAGCAACAACAACAACAACAGCA 289391

Qy        7821 AAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACC 7880
               |  |      |  | ||     || |  || ||||| || || || | ||||||   ||
Db      289392 ACAACAACAGCAGCAGCAACAACAGCTACAACAGCAACAACAACAACAACAGCAATTACA 289451

Qy        7881 TCGAGTGTATGAGTGACGAGTACAACATGAACCG 7914
                | |  | |  ||  || |  ||||||  ||| |
Db      289452 ACAACAGCAGCAGCAACAACAACAACAACAACAG 289485


RESULT 15
AC130618/c
LOCUS       AC130618               211295 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-198A11, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC130618
VERSION     AC130618.4  GI:30580715
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 211295)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 211295)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 211295)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23268237.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
```

```
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented, and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GIYR
              Center clone name: CH230-198A11
          -------------- Summary Statistics
              Assembly program: Atlas 3.0;
              Consensus quality: 189644 bases at least Q40
              Consensus quality: 193256 bases at least Q30
              Consensus quality: 195687 bases at least Q20
              Estimated insert size: 198972; sum-of-contigs estimation
              Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 3 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1   179760: contig of 179760 bp in length
          *   179761   179860: gap of unknown length
          *   179861   209996: contig of 30136 bp in length
          *   209997   210096: gap of unknown length
          *   210097   211295: contig of 1199 bp in length.
FEATURES             Location/Qualifiers
     source          1..211295
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-198A11"
     misc_feature    complement(1563..2142)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH345664"
     misc_feature    complement(178560..179255)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH345665"
     misc_feature    179861..180867
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             3.3%;  Score 267.8;  DB 2;  Length 211295;
  Best Local Similarity   69.8%;  Pred. No. 3e-37;
  Matches  376;  Conservative    0;  Mismatches  162;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db      29573 ACAACAGCAACAACAACAGCAACAACAACAGCAACAACAACAGCAACAGCAACAACAACA 29514

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db      29513 GCAACAACAACAGCAACAACAACAACAGCAACAACAACAGCAGCAACAGCAACAACAACA 29454

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db      29453 GCAACAACAACAGCAGCAACAACAACAACAACAGCAACAACAACAGCAGCAACAACAACA 29394

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db      29393 ACAGCAACAGCAACAACAACAGCAACAACAACAACAGCAACAACAACAACAACAGCAACA 29334

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGA-CAACAACA 7700
Db      29333 ACAACAGCAGCAACAACAACAGCAACTACAACAGCAACAACAACAACAGCATCAACAACA 29274

Qy       7701 AGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTT 7760
Db      29273 ACAACAACAGCTACAACAACAACTGCAACATCAACAGCAACAACAACAACAGCAGCAACA 29214

Qy       7761 ATGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCA 7820
Db      29213 ACAGCAACAACAACAGCAACAACAACAACAGCAACAGCAACAACAACAGCAGCAACAACA 29154

Qy       7821 AAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACC 7880
Db      29153 ACAACAGCAACAACAACAACAGCAACAACAACAGCAACAACAACAGCAACAACAACAACA 29094

Qy       7881 TCGAGTGTATGAGTGACGAGTACAACATGAACCGTCGACGATCCCTATAGTGAGTCGTA 7939
Db      29093 ACAACAGCAGCAACAACAACAGCAACAACAACAATCTACTATCATTTAATTTAGTAATA 29035

Search completed: March 30, 2004, 22:38:42
Job time : 19856 secs
```

```
OM nucleic - nucleic search, using sw model

Run on:          March 30, 2004, 15:31:34 ; Search time 1840 Seconds
                                            (without alignments)
                                            18574.323 Million cell updates/sec

Title:           US-09-913-878A-1
Perfect score:   8045
Sequence:        1 ggtaccgggcccccctcga..........cgtattacaatcccagtcac 8045

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:      6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        N_Geneseq_29Jan04:*
                 1:  geneseqn1980s:*
                 2:  geneseqn1990s:*
                 3:  geneseqn2000s:*
                 4:  geneseqn2001as:*
                 5:  geneseqn2001bs:*
                 6:  geneseqn2002s:*
                 7:  geneseqn2003as:*
                 8:  geneseqn2003bs:*
                 9:  geneseqn2003cs:*
                 10:  geneseqn2004s:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	8045	100.0	8045	3	AAA65171	Aaa65171 Neurospor
	2	220.8	2.7	5163	2	AAV20700	Aav20700 Cryptospo
	3	220.8	2.7	5163	3	AAA61849	Aaa61849 ORF encod
	4	220.8	2.7	5163	6	ABT04778	Abt04778 C parvum
	5	220.8	2.7	5318	3	AAA61848	Aaa61848 DNA encod
	6	220.8	2.7	5318	6	ABT04777	Abt04777 C parvum
	7	219.2	2.7	5318	2	AAV20701	Aav20701 Cryptospo
	8	211.4	2.6	1236	6	AAD44410	Aad44410 Human hun
c	9	205	2.5	567	3	AAA29550	Aaa29550 HIV codon
c	10	203.8	2.5	2336	4	ABL25662	Abl25662 Drosophil
	11	200.4	2.5	5511	3	AAA61847	Aaa61847 Cryptospo
	12	200.4	2.5	5511	6	ABT04776	Abt04776 C parvum
	13	200.4	2.5	7334	3	AAA61846	Aaa61846 Cryptospo
	14	200.4	2.5	7334	6	ABT04775	Abt04775 C parvum
	15	196.6	2.4	486	4	AAF75507	Aaf75507 Polygluta
	16	194.6	2.4	867	2	AAX91461	Aax91461 T. gondii
	17	194.6	2.4	867	2	AAX91460	Aax91460 T. gondii
	18	194.6	2.4	867	4	AAS42783	Aas42783 T. gondii
	19	194.6	2.4	867	4	AAS42784	Aas42784 Reverse c
	20	194.6	2.4	1397	2	AAX91462	Aax91462 T. gondii
c	21	194.6	2.4	1397	2	AAX91463	Aax91463 T. gondii
c	22	194.6	2.4	1397	4	AAS42786	Aas42786 Reverse c
	23	194.6	2.4	1397	4	AAS42785	Aas42785 T. gondii
	24	182.8	2.3	13631	4	ABL20354	Abl20354 Drosophil
c	25	177	2.2	7758	6	ABL33103	Abl33103 Human imm
	26	176.6	2.2	198	2	AAT17204	Aat17204 DNA-spann
	27	176.6	2.2	198	2	AAT17205	Aat17205 DNA-spann
	28	175	2.2	198	2	AAT17202	Aat17202 DNA-spann
	29	175	2.2	198	2	AAT17203	Aat17203 DNA-spann
c	30	171.8	2.1	1400	5	AAS83195	Aas83195 DNA encod
	31	162.8	2.0	4829	4	ABL13399	Abl13399 Drosophil
c	32	162.8	2.0	39746	4	ABL13398	Abl13398 Drosophil
	33	160	2.0	2717	6	ABQ76406	Abq76406 S. cerevi
	34	158.8	2.0	2000	7	ACC61172	Acc61172 Gene sequ
	35	158.8	2.0	3218	6	ABQ76319	Abq76319 S. cerevi
	36	157.4	2.0	1972	9	ADC87176	Adc87176 Human GPC
	37	157.4	2.0	2032	2	AAT10117	Aat10117 Drosophil
	38	157.4	2.0	2032	4	ABL05381	Abl05381 Drosophil
	39	157.4	2.0	2032	6	ABS53745	Abs53745 cDNA enco
	40	157.4	2.0	2032	6	AAD33916	Aad33916 Drosophil
c	41	157.4	2.0	4282	4	ABL05380	Abl05380 Drosophil
c	42	157.4	2.0	6668	6	ABL33697	Abl33697 Human imm
	43	156.8	1.9	6948	4	AAH62808	Aah62808 Shrimp wh
	44	156.8	1.9	305107	4	AAH62689	Aah62689 Shrimp wh
	45	156.4	1.9	2550	4	AAS23437	Aas23437 Candida a

ALIGNMENTS

```
RESULT 1
AAA65171
ID   AAA65171 standard; DNA; 8045 BP.
XX
AC   AAA65171;
XX
DT   16-NOV-2000  (first entry)
XX
DE   Neurospora crassa qde-1 gene.
XX
KW   Gene silencing; quelling deficient; qde-1; al-1; ds.
XX
OS   Neurospora crassa.
XX
FH   Key             Location/Qualifiers
FT   CDS             2447. .6655
FT                   /*tag=  a
FT                   /product= "QDE-1"
XX
PN   WO200050581-A2.
XX
PD   31-AUG-2000.
XX
PF   16-FEB-2000; 2000WO-IT000048.
XX
PR   22-FEB-1999;   99IT-RM000117.
XX
PA   (UYRO-) UNIV ROMA LA SAPIENZA.
XX
PI   Macino G,  Cogoni C;
XX
DR   WPI; 2000-579171/54.
DR   P-PSDB; AAB13956.
XX
PT   Novel polynucleotide encoding a polypeptide which has a silencing
PT   activity and comprising a RNA-dependent RNA polymerase domain.
XX
PS   Claim 1; Page 31-43; 48pp; English.
XX
CC   The present sequence is the Neurospora crassa qde-1 gene. This gene has
CC   silencing activity. The qde-1 gene was isolated by mutational analysis of
CC   an al-1 transgenic strain. This strain had an albino phenotype resulting
CC   from post-transcriptional silencing of the endogenous al-1 gene.
CC   Reversion of this phenotype indicated a mutation in a silencing gene. The
CC   silencing gene, qde-1, could then be isolated. Modulation of qde-1
CC   expression may be used to inactivate genes and to silence suppression of
```

```
CC   genes
XX
SQ   Sequence 8045 BP; 2089 A; 2180 C; 1950 G; 1826 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 8045;  DB 3;  Length 8045;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 8045;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCTGAAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCTGAAAA 60

Qy         61 GTCCTAAAGTCTCCTGTCTCTCATGCTAGGTAGCCCTACTGCGTGTGTAGTGTAGTAGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTCCTAAAGTCTCCTGTCTCTCATGCTAGGTAGCCCTACTGCGTGTGTAGTGTAGTAGGT 120

Qy        121 CCGAATAGGTTCGTACGCTGTCCCCTTTAGGGGGCTGGTTTAACAGGGAGACAGGGTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGAATAGGTTCGTACGCTGTCCCCTTTAGGGGGCTGGTTTAACAGGGAGACAGGGTACA 180

Qy        181 GTCTCTCGGGAACAGAGACGACCAAGCTTTGTTGTGTGACATCATCTTGAGACCATTCGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCTCTCGGGAACAGAGACGACCAAGCTTTGTTGTGTGACATCATCTTGAGACCATTCGT 240

Qy        241 CTGAGGGTCCATACAACCATACATGCTGTGTAGAGAGTGGTGGGAATATTAGAGGTTACT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGAGGGTCCATACAACCATACATGCTGTGTAGAGAGTGGTGGGAATATTAGAGGTTACT 300

Qy        301 TTGAGAGTTTCTGAAGAGATTGAGACTGTAGGTACGTTCATTTATCGGTGCGACGAGAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGAGAGTTTCTGAAGAGATTGAGACTGTAGGTACGTTCATTTATCGGTGCGACGAGAAT 360

Qy        361 GGGCACGCATGAGTCGACATTTGCAATTTGTTGGTTGGCACGCACATCGATAGTGTACAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGGCACGCATGAGTCGACATTTGCAATTTGTTGGTTGGCACGCACATCGATAGTGTACAC 420

Qy        421 ATGTGTCGGCCTGAGACTGTCCAATGGACCAATGGTAAGGCGTGAGCGGCGAGTGGGGTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATGTGTCGGCCTGAGACTGTCCAATGGACCAATGGTAAGGCGTGAGCGGCGAGTGGGGTC 480

Qy        481 AACTTGAGGTTAATGTACACAGTACTTGAGGCTGTAGCGGGCAGCGGAAGCATCCGTCCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AACTTGAGGTTAATGTACACAGTACTTGAGGCTGTAGCGGGCAGCGGAAGCATCCGTCCT 540

Qy        541 CGAATGGACATTCCCAAGCCAGCCAACACTGGTGTCCATTTAAAAGCGTTTTCTCCATGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGAATGGACATTCCCAAGCCAGCCAACACTGGTGTCCATTTAAAAGCGTTTTCTCCATGG 600

Qy        601 GGGAACAATGGCCGGCGTAGTATCTGTAAGAGGACCTAGAGGACGGTATTAGAACATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGGAACAATGGCCGGCGTAGTATCTGTAAGAGGACCTAGAGGACGGTATTAGAACATTTG 660

Qy        661 GGTGATGTATCTAGATCACCGATGTGTCCCTGAATAGTCCATGAATAGTCTGGGATGAGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGTGATGTATCTAGATCACCGATGTGTCCCTGAATAGTCCATGAATAGTCTGGGATGAGT 720

Qy        721 CCGAGATCAGTCCGGGATGTGTCACCGATATATCCATGACTTATCTCTGATTTATCCCAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCGAGATCAGTCCGGGATGTGTCACCGATATATCCATGACTTATCTCTGATTTATCCCAG 780

Qy        781 CGAGACCCCAATCAGTATGACCGACAGCGACACCATCACCGGCGCCAGCGTATGAGATCC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGAGACCCCAATCAGTATGACCGACAGCGACACCATCACCGGCGCCAGCGTATGAGATCC 840

Qy        841 TGAACATATGTTAGGCATATGTATGGCATATGTATGCATAGAGCATGCACCACATATCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGAACATATGTTAGGCATATGTATGGCATATGTATGCATAGAGCATGCACCACATATCTG 900

Qy        901 AGAGCTGCCTGTGACTACCTAGAGGTATGTAGAGGTAATGTCACGTACCTTACCTCTACC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGAGCTGCCTGTGACTACCTAGAGGTATGTAGAGGTAATGTCACGTACCTTACCTCTACC 960

Qy        961 AAATGTATCCCGCAGTCACTCGTATTCCCGCTGTAGTGTACACTAGAACTAGCCCCGCCT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAATGTATCCCGCAGTCACTCGTATTCCCGCTGTAGTGTACACTAGAACTAGCCCCGCCT 1020

Qy       1021 TTTGGATCCTGGAGTGTCCCCATTCCACGCACTTGCCAAGAACCACAACCTCGCAAAGCG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTTGGATCCTGGAGTGTCCCCATTCCACGCACTTGCCAAGAACCACAACCTCGCAAAGCG 1080

Qy       1081 CGCTGACAGGGCACCGTGCTGCCTGGACCTTTGCGGGCCGTTCCTTCTTGACCGGCGCCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCTGACAGGGCACCGTGCTGCCTGGACCTTTGCGGGCCGTTCCTTCTTGACCGGCGCCA 1140

Qy       1141 GGCAGCAGACCGGTTGGTTCTTGTTCCTCAAGGTGACGAAGGCGATTAAAAAATCCAAGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GGCAGCAGACCGGTTGGTTCTTGTTCCTCAAGGTGACGAAGGCGATTAAAAAATCCAAGA 1200

Qy       1201 CAACCAGATCCTGGAACCTTTGATCATCCCTCGTTTACTCAGACGAAAAAACTAAAAAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CAACCAGATCCTGGAACCTTTGATCATCCCTCGTTTACTCAGACGAAAAAACTAAAAAAC 1260

Qy       1261 TTCAAAGTCGTTGTCAGATTAATTGAACCGTTAAAACAGGACATGTCCGCCCTTTCGTCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCAAAGTCGTTGTCAGATTAATTGAACCGTTAAAACAGGACATGTCCGCCCTTTCGTCT 1320

Qy       1321 CTCATGCCCTCGAAACCTCACGTCTTCCATCTCTCGTCCGATCCCCCCCTCCCCACGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCATGCCCTCGAAACCTCACGTCTTCCATCTCTCGTCCGATCCCCCCCTCCCCACGACC 1380

Qy       1381 CTGATTCAGCAGCGAACTGACCTGTCACTTTACTGTGGACTCCTCCTAGAGTAGGGCTGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTGATTCAGCAGCGAACTGACCTGTCACTTTACTGTGGACTCCTCCTAGAGTAGGGCTGG 1440

Qy       1441 CCATCCCCGGGCCCCCTGCAGCTTCTGCAAGTCAAAGCGATGGAAGGATGGCCGTCCAGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCATCCCCGGGCCCCCTGCAGCTTCTGCAAGTCAAAGCGATGGAAGGATGGCCGTCCAGT 1500

Qy       1501 GGGTGCGGCATACAGCCATTTCCAAAATGGTCTCGTCGCATGCGGAAGAAAAAAAATGGA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GGGTGCGGCATACAGCCATTTCCAAAATGGTCTCGTCGCATGCGGAAGAAAAAAAATGGA 1560

Qy       1561 AGGCTACGAGCCTGCGACACCTGTACGGTCACTGCGGCTACCCTTCACCACCACCAGTGA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 AGGCTACGAGCCTGCGACACCTGTACGGTCACTGCGGCTACCCTTCACCACCACCAGTGA 1620

Qy       1621 TATCATCACCCATACCATCAGGGCCACTACCGCCGCGCTCTCTCGAACATCGGGCTGAAT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TATCATCACCCATACCATCAGGGCCACTACCGCCGCGCTCTCTCGAACATCGGGCTGAAT 1680

Qy       1681 GTGACCAATTGCGCCGACAGCGGGCTTTTCGATCCTGTGGGCAGCTTCTTCCACCGGTTT 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGACCAATTGCGCCGACAGCGGGCTTTTCGATCCTGTGGGCAGCTTCTTCCACCGGTTT 1740

Qy       1741 GCGCTGATCTGGGAGCGTATCGAGCTGGATTGGGACTGTTTACCACTCGAATCACTTGAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GCGCTGATCTGGGAGCGTATCGAGCTGGATTGGGACTGTTTACCACTCGAATCACTTGAC 1800

Qy       1801 GGCTCGGCGCGATTCTATGGCCATCCACCGCGATTGATCGGTCCCTTCACTGTTTCCTGC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GGCTCGGCGCGATTCTATGGCCATCCACCGCGATTGATCGGTCCCTTCACTGTTTCCTGC 1860

Qy       1861 AACCTACTCTCCCGACCAGATCAGACGAGGCTTTATGCCCAGTTCCGGGCGTTGCCAATT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AACCTACTCTCCCGACCAGATCAGACGAGGCTTTATGCCCAGTTCCGGGCGTTGCCAATT 1920

Qy       1921 ATAGCCACAGCACGAAGCTCGCCAAGTTCGTTCGAAATGCTCAAAACTGTTCACACACTT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ATAGCCACAGCACGAAGCTCGCCAAGTTCGTTCGAAATGCTCAAAACTGTTCACACACTT 1980

Qy       1981 TGCACGAACAGGACCACCACGAAAAACCACCTTGGCCCGCTCGCGCGAACCCGTGTCTTT 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TGCACGAACAGGACCACCACGAAAAACCACCTTGGCCCGCTCGCGCGAACCCGTGTCTTT 2040

Qy       2041 GAACTCGATCAGCCCCAATCCCAAACCAGGTCGCCTTTCCCCGCTGCGCCTGCTTGTCCC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  2041 GAACTCGATCAGCCCCAATCCCAAACCAGGTCGCCTTTCCCCGCTGCGCCTGCTTGTCCC 2100

Qy  2101 GACTTCGTGAAGCAAACAAGACTGGCGGCTATGGCCAAGGCCCATTCCATGACAATTGGT 2160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 GACTTCGTGAAGCAAACAAGACTGGCGGCTATGGCCAAGGCCCATTCCATGACAATTGGT 2160

Qy  2161 TAATCGAGGTCAAAAACTACAAGAGATCAACGCCGCAAAGATATACTTGTCGTTCACCCC 2220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 TAATCGAGGTCAAAAACTACAAGAGATCAACGCCGCAAAGATATACTTGTCGTTCACCCC 2220

Qy  2221 GACAACCGCAAAGCCTTGAAGCGCAAAGCCACAGTGGTGTTGTTTACTTCCGCACACTTA 2280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 GACAACCGCAAAGCCTTGAAGCGCAAAGCCACAGTGGTGTTGTTTACTTCCGCACACTTA 2280

Qy  2281 TCGACCCTTACGCCCAAGACGCTCCCGCTTGGTAGAAGCTCGAGTCGGGTAATCAGGTAT 2340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 TCGACCCTTACGCCCAAGACGCTCCCGCTTGGTAGAAGCTCGAGTCGGGTAATCAGGTAT 2340

Qy  2341 ATCTCACAGTCGCCTTGTCCTCCTTTTCGGTGGTGGTCGCTTTTATCATGCCTTATCGCT 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 ATCTCACAGTCGCCTTGTCCTCCTTTTCGGTGGTGGTCGCTTTTATCATGCCTTATCGCT 2400

Qy  2401 AACCCGGTTGTATACTCTACTCCAGGCCCGATCTCCCACGGCACAAATGAACCCTATTAC 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 AACCCGGTTGTATACTCTACTCCAGGCCCGATCTCCCACGGCACAAATGAACCCTATTAC 2460

Qy  2461 TCCTAGGAAGAGGAATAGCCCCGTCGAGGAAATCATAAACCGGCTCAATAACGACTACAA 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 TCCTAGGAAGAGGAATAGCCCCGTCGAGGAAATCATAAACCGGCTCAATAACGACTACAA 2520

Qy  2521 CCTGGGCCTCCAGTGTGTCGCAGACACAACTCTCACCCCCCACCGCCGGAAGGAGCTGGC 2580
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 CCTGGGCCTCCAGTGTGTCGCAGACACAACTCTCACCCCCCACCGCCGGAAGGAGCTGGC 2580

Qy  2581 CGAGAGTGACGAGGATTTCGGTCGCCATGACAAGATCTACAGAGCCCTGAACTTTCTCTA 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581 CGAGAGTGACGAGGATTTCGGTCGCCATGACAAGATCTACAGAGCCCTGAACTTTCTCTA 2640

Qy  2641 CTGGCGGAAGGATGACTCCCTGAACCAGGCAGAAGCCAACTTCTTCATCGAGGCCAAAGC 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641 CTGGCGGAAGGATGACTCCCTGAACCAGGCAGAAGCCAACTTCTTCATCGAGGCCAAAGC 2700

Qy  2701 TGCGAGCTCGAACTGGGTGCCCAAAGCCCACGCCGACCCTGACACGCTTCCGTGGTCCAA 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701 TGCGAGCTCGAACTGGGTGCCCAAAGCCCACGCCGACCCTGACACGCTTCCGTGGTCCAA 2760

Qy  2761 GGAACCTCCCCGCGCCGCTACTGCCGGCCAACAATGGGCATTGCAGACTGTGTTGCTCGA 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761 GGAACCTCCCCGCGCCGCTACTGCCGGCCAACAATGGGCATTGCAGACTGTGTTGCTCGA 2820

Qy  2821 GGTGCTTAATAGGTTTATGCCACCTCCCAATAACACACCAGGTCGAACGTTTGGCAGAAC 2880
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821 GGTGCTTAATAGGTTTATGCCACCTCCCAATAACACACCAGGTCGAACGTTTGGCAGAAC 2880

Qy  2881 TCTAAGCGGCCCAAGTGGCCTGAGCCGCCCAACCTCTACCAACACCAAACGCAAGGATGA 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881 TCTAAGCGGCCCAAGTGGCCTGAGCCGCCCAACCTCTACCAACACCAAACGCAAGGATGA 2940

Qy  2941 GCCCGCCAATGTCACTTTCGCTGATCCGCCCAAACGCTCGTTGACTCGCTCTGCCACAGG 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941 GCCCGCCAATGTCACTTTCGCTGATCCGCCCAAACGCTCGTTGACTCGCTCTGCCACAGG 3000

Qy  3001 TCCTCCTATTCACGGCGCGGCGATACCCCTAAAGTTCCCCGATCCAGTGAATACCGGTTC 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001 TCCTCCTATTCACGGCGCGGCGATACCCCTAAAGTTCCCCGATCCAGTGAATACCGGTTC 3060

Qy  3061 CAAACGACCATCTCTCGAGAGTGAGAATCTCAATCAGTGCACCAAGCGGGCCAAGGGCAA 3120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061 CAAACGACCATCTCTCGAGAGTGAGAATCTCAATCAGTGCACCAAGCGGGCCAAGGGCAA 3120

Qy  3121 GCTGTCTGATAATGTTGCCGCTGCCGCCGCCCCGCCCGTGCCTATTGCGAGCGCTTTGGA 3180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121 GCTGTCTGATAATGTTGCCGCTGCCGCCGCCCCGCCCGTGCCTATTGCGAGCGCTTTGGA 3180

Qy  3181 CAAGGTACCGACTCGAAGGCATGCCAATACGAGAGATCCCACGGCGACAGGTCATAGACG 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181 CAAGGTACCGACTCGAAGGCATGCCAATACGAGAGATCCCACGGCGACAGGTCATAGACG 3240

Qy  3241 AGCGGACCAGGTGGATTCCTTTGATACATCTCAAGGCACTTCCTATGGTTCGAGTGTCTT 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241 AGCGGACCAGGTGGATTCCTTTGATACATCTCAAGGCACTTCCTATGGTTCGAGTGTCTT 3300

Qy  3301 CAGCGCTTGCCGTCACAATCAGAGCACTACCCAGAGTAGTTTTGAGGCTCCTCCTTCACA 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301 CAGCGCTTGCCGTCACAATCAGAGCACTACCCAGAGTAGTTTTGAGGCTCCTCCTTCACA 3360

Qy  3361 GCCCAGAGAGAAGCGGCCTGTGGATGCCACGGTCTTTGAGGCTGGACACTTGATTGAGTC 3420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361 GCCCAGAGAGAAGCGGCCTGTGGATGCCACGGTCTTTGAGGCTGGACACTTGATTGAGTC 3420

Qy  3421 TCCTAGCAAAGGAAGAACAACCAAGTCCCACATAGATAACCAGCCCCTTTCATCGTCTTC 3480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3421 TCCTAGCAAAGGAAGAACAACCAAGTCCCACATAGATAACCAGCCCCTTTCATCGTCTTC 3480

Qy  3481 CCAGGGTGAAACTTCGTTCAGCACTTACTATGAGTCGTTTCCAAGTTCCGGCGGCGAGGG 3540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3481 CCAGGGTGAAACTTCGTTCAGCACTTACTATGAGTCGTTTCCAAGTTCCGGCGGCGAGGG 3540

Qy  3541 CGCAATTCCCGAGCCGAGTCGCTCAAATGGACTGGCTCGGAGCGAAGAAAGCGCTCGATC 3600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3541 CGCAATTCCCGAGCCGAGTCGCTCAAATGGACTGGCTCGGAGCGAAGAAAGCGCTCGATC 3600

Qy  3601 TCAGGTTCAAGTTCATGCCCCGGTGGTTGCAGCTCGGCTGAGAAATATTTGGCCGAAATT 3660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3601 TCAGGTTCAAGTTCATGCCCCGGTGGTTGCAGCTCGGCTGAGAAATATTTGGCCGAAATT 3660

Qy  3661 TCCCAAATGGCTACACGAAGCTCCTCTCGCTGTTGCATGGGAAGTTACCAGACTCTTTAT 3720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3661 TCCCAAATGGCTACACGAAGCTCCTCTCGCTGTTGCATGGGAAGTTACCAGACTCTTTAT 3720

Qy  3721 GCACTGCAAAGTAGACTTGGAAGACGAGAGCCTGGGCCTAAAGTACGACCCTTCCTGGTC 3780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3721 GCACTGCAAAGTAGACTTGGAAGACGAGAGCCTGGGCCTAAAGTACGACCCTTCCTGGTC 3780

Qy  3781 TACCGCGCGCGATGTCACAGATATCTGGAAGACTCTCTACCGGCTTGATGCTTTCCGTGG 3840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3781 TACCGCGCGCGATGTCACAGATATCTGGAAGACTCTCTACCGGCTTGATGCTTTCCGTGG 3840

Qy  3841 TAAACCCTTTCCAGAAAAGCCGCCCAACGACGTGTTCGTGACCGCAATGACGGGCAACTT 3900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3841 TAAACCCTTTCCAGAAAAGCCGCCCAACGACGTGTTCGTGACCGCAATGACGGGCAACTT 3900

Qy  3901 TGAGAGCAAAGGTAGTGCCGTTGTTCTCTCTGCTGTTCTAGACTACAATCCGGACAACTC 3960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3901 TGAGAGCAAAGGTAGTGCCGTTGTTCTCTCTGCTGTTCTAGACTACAATCCGGACAACTC 3960

Qy  3961 GCCTACTGCGCCCCTTTACCTTGTGAAGCTGAAGCCGCTCATGTTCGAGCAGGGCTGTCG 4020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3961 GCCTACTGCGCCCCTTTACCTTGTGAAGCTGAAGCCGCTCATGTTCGAGCAGGGCTGTCG 4020

Qy  4021 ACTCACCCGTCGGTTCGGTCCTGATAGGTTTTTCGAGATCCTTATACCGTCGCCTACGAG 4080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4021 ACTCACCCGTCGGTTCGGTCCTGATAGGTTTTTCGAGATCCTTATACCGTCGCCTACGAG 4080

Qy  4081 CACCAGCCCAAGTGTACCGCCGGTGGTCAGCAAACAACCAGGTGCGGTCGAAGAAGTCAT 4140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4081 CACCAGCCCAAGTGTACCGCCGGTGGTCAGCAAACAACCAGGTGCGGTCGAAGAAGTCAT 4140

Qy  4141 CCAGTGGCTCACGATGGGGCAACATTCTCTTGTAGGCCGCCAATGGCGCGCTTTCTTCGC 4200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4141 CCAGTGGCTCACGATGGGGCAACATTCTCTTGTAGGCCGCCAATGGCGCGCTTTCTTCGC 4200

Qy  4201 CAAAGATGCCGGATACAGGAAACCTCTCAGGGAGTTCCAGCTCCGCGCCGAGGACCCGAA 4260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4201 CAAAGATGCCGGATACAGGAAACCTCTCAGGGAGTTCCAGCTCCGCGCCGAGGACCCGAA 4260
```

```
Qy     4261 ACCCATCATCAAGGAGAGAGTCCACTTCTTTGCCGAGACCGGCATTACGTTCCGACCTGA 4320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4261 ACCCATCATCAAGGAGAGAGTCCACTTCTTTGCCGAGACCGGCATTACGTTCCGACCTGA 4320

Qy     4321 TGTGTTCAAGACGAGATCTGTCGTTCCGGCAGAGGAACCTGTAGAGCAACGGACCGAGTT 4380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4321 TGTGTTCAAGACGAGATCTGTCGTTCCGGCAGAGGAACCTGTAGAGCAACGGACCGAGTT 4380

Qy     4381 CAAAGTTAGTCAAATGCTGGACTGGCTCCTGCAACTCGACAACAACACTTGGCAGCCGCA 4440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4381 CAAAGTTAGTCAAATGCTGGACTGGCTCCTGCAACTCGACAACAACACTTGGCAGCCGCA 4440

Qy     4441 CCTCAAGTTGTTCTCCCGTATCCAGCTCGGTCTGAGTAAGACATATGCCATTATGACATT 4500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4441 CCTCAAGTTGTTCTCCCGTATCCAGCTCGGTCTGAGTAAGACATATGCCATTATGACATT 4500

Qy     4501 GGAGCCTCACCAGATCAGACACCACAAGACCGATCTTCTTTCACCTTCAGGCACTGGCGA 4560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4501 GGAGCCTCACCAGATCAGACACCACAAGACCGATCTTCTTTCACCTTCAGGCACTGGCGA 4560

Qy     4561 AGTGATGAATGACGGTGTAGGCCGCATGTCGCGAAGCGTGGCCAAGAGGATACGCGATGT 4620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4561 AGTGATGAATGACGGTGTAGGCCGCATGTCGCGAAGCGTGGCCAAGAGGATACGCGATGT 4620

Qy     4621 TCTCGGTTTGGGTGATGTGCCCTCTGCTGTGCAAGGGCGGTTTGGTTCGGCCAAGGGAAT 4680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4621 TCTCGGTTTGGGTGATGTGCCCTCTGCTGTGCAAGGGCGGTTTGGTTCGGCCAAGGGAAT 4680

Qy     4681 GTGGGTTATCGACGTTGACGACACAGGCGATGAGGATTGGATCGAGACATACCCGTCCCA 4740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4681 GTGGGTTATCGACGTTGACGACACAGGCGATGAGGATTGGATCGAGACATACCCGTCCCA 4740

Qy     4741 GCGCAAGTGGGAATGCGACTTCGTTGATAAACATCAACGTACCCTCGAAGTCCGGAGCGT 4800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4741 GCGCAAGTGGGAATGCGACTTCGTTGATAAACATCAACGTACCCTCGAAGTCCGGAGCGT 4800

Qy     4801 GGCTTCTGAACTGAAGTCAGCTGGTCTCAACCTACAGCTGTTACCTGTCCTGGAAGATAG 4860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4801 GGCTTCTGAACTGAAGTCAGCTGGTCTCAACCTACAGCTGTTACCTGTCCTGGAAGATAG 4860

Qy     4861 AGCCAGGGACAAGGTGAAGATGCGCCAGGCAATCGGTGACCGTCTTATCAACGATTTGCA 4920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4861 AGCCAGGGACAAGGTGAAGATGCGCCAGGCAATCGGTGACCGTCTTATCAACGATTTGCA 4920

Qy     4921 ACGACAGTTCAGCGAGCAAAAGCATGCTTTGAATCGCCCAGTGGAATTTCGCCAATGGGT 4980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4921 ACGACAGTTCAGCGAGCAAAAGCATGCTTTGAATCGCCCAGTGGAATTTCGCCAATGGGT 4980

Qy     4981 TTACGAGAGTTATTCCAGTCGCGCAACTCGAGTCAGCCACGGCCGTGTGCCTTTTCTTGC 5040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4981 TTACGAGAGTTATTCCAGTCGCGCAACTCGAGTCAGCCACGGCCGTGTGCCTTTTCTTGC 5040

Qy     5041 TGGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAACAGTGGGTTCGATCC 5100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5041 TGGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAACAGTGGGTTCGATCC 5100

Qy     5101 CAAGAAGCAAAAGTACTTGCAAGACATCGCCTGGGATCTTCAAAAGCGGAAATGTGACAC 5160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5101 CAAGAAGCAAAAGTACTTGCAAGACATCGCCTGGGATCTTCAAAAGCGGAAATGTGACAC 5160

Qy     5161 GTTGAAGTCCAAGCTGAACATCCGTGTCGGTCGATCAGCATACATTTACATGATTGCCGA 5220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5161 GTTGAAGTCCAAGCTGAACATCCGTGTCGGTCGATCAGCATACATTTACATGATTGCCGA 5220

Qy     5221 TTTCTGGGGTGTGCTTGAGGAAAATGAGGTTCATGTCGGATTCTCCTCAAAGTTCAGAGA 5280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5221 TTTCTGGGGTGTGCTTGAGGAAAATGAGGTTCATGTCGGATTCTCCTCAAAGTTCAGAGA 5280

Qy     5281 CGAGGAGGAGTCTTTTACACTCCTATCGGACTGTGATGTCCTCGTGGCGCGATCCCCAGC 5340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5281 CGAGGAGGAGTCTTTTACACTCCTATCGGACTGTGATGTCCTCGTGGCGCGATCCCCAGC 5340

Qy     5341 CCATTTCCCTAGTGATATCCAACGGGTTCGAGCAGTCTTCAAGCCAGAGCTCCACAGTCT 5400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5341 CCATTTCCCTAGTGATATCCAACGGGTTCGAGCAGTCTTCAAGCCAGAGCTCCACAGTCT 5400

Qy     5401 CAAGGATGTAATCATCTTCTCTACTAAAGGAGATGTACCGCTTGCTAAGAAGCTATCTGG 5460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5401 CAAGGATGTAATCATCTTCTCTACTAAAGGAGATGTACCGCTTGCTAAGAAGCTATCTGG 5460

Qy     5461 TGGAGACTACGACGGCGATATGGCCTGGGTCTGCTGGGATCCGGAGATCGTCGATGGTTT 5520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5461 TGGAGACTACGACGGCGATATGGCCTGGGTCTGCTGGGATCCGGAGATCGTCGATGGTTT 5520

Qy     5521 CGTCAATGCGGAAATGCCTCTGGAGCCCGACCTGTCTAGGTACCTAAAGAAGGACAAAAC 5580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5521 CGTCAATGCGGAAATGCCTCTGGAGCCCGACCTGTCTAGGTACCTAAAGAAGGACAAAAC 5580

Qy     5581 GACTTTCAAACAACTTATGGCCTCACACGGCACGGGCTCAGCGGCCAAAGAGCAGACTAC 5640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5581 GACTTTCAAACAACTTATGGCCTCACACGGCACGGGCTCAGCGGCCAAAGAGCAGACTAC 5640

Qy     5641 ATACGATATGATCCAGAAGAGCTTCCATTTCGCCCTGCAGCCCAACTTCTTGGGCATGTG 5700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5641 ATACGATATGATCCAGAAGAGCTTCCATTTCGCCCTGCAGCCCAACTTCTTGGGCATGTG 5700

Qy     5701 CACTAACTACAAAGAAAGGCTCTGTTACATCAACAATAGTGTGTCTAACAAGCCGGCCAT 5760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5701 CACTAACTACAAAGAAAGGCTCTGTTACATCAACAATAGTGTGTCTAACAAGCCGGCCAT 5760

Qy     5761 CATTCTTAGTTCACTGGTGGGAAACCTCGTCGATCAGAGCAAGCAAGGTATTGTCTTTAA 5820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5761 CATTCTTAGTTCACTGGTGGGAAACCTCGTCGATCAGAGCAAGCAAGGTATTGTCTTTAA 5820

Qy     5821 CGAAGCAAGCTGGGCTCAATTGCGTAGGGAACTGCTTGGCGGTGCATTGTCCCTTCCTGA 5880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5821 CGAAGCAAGCTGGGCTCAATTGCGTAGGGAACTGCTTGGCGGTGCATTGTCCCTTCCTGA 5880

Qy     5881 CCCAATGTACAAGAGCGACAGTTGGCTCGGGCGCGGAGAGCCTACCCACATTATTGACTA 5940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5881 CCCAATGTACAAGAGCGACAGTTGGCTCGGGCGCGGAGAGCCTACCCACATTATTGACTA 5940

Qy     5941 CCTGAAATTCTCCATCGCCAGGCCTGCGATTGACAAGGAACTGGAAGCCTTCCACAATGC 6000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5941 CCTGAAATTCTCCATCGCCAGGCCTGCGATTGACAAGGAACTGGAAGCCTTCCACAATGC 6000

Qy     6001 CATGAAAGCGGCCAAGGATACAGAAGACGGCGCTCACTTTTGGGATCCGGATCTCGCTTC 6060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6001 CATGAAAGCGGCCAAGGATACAGAAGACGGCGCTCACTTTTGGGATCCGGATCTCGCTTC 6060

Qy     6061 CTACTACACGTTCTTCAAGGAGATTAGCGACAAGTCGCGATCGTCCGCACTGCTATTCAC 6120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6061 CTACTACACGTTCTTCAAGGAGATTAGCGACAAGTCGCGATCGTCCGCACTGCTATTCAC 6120

Qy     6121 GACTCTGAAGAACCGTATCGGCGAAGTCGAGAAAGAATATGGCAGGTTGGTCAAAAACAA 6180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6121 GACTCTGAAGAACCGTATCGGCGAAGTCGAGAAAGAATATGGCAGGTTGGTCAAAAACAA 6180

Qy     6181 GGAGATGAGAGACAGCAAGGACCCCTACCCTGTCCGCGTCAACCAGGTTTATGAAAAATG 6240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6181 GGAGATGAGAGACAGCAAGGACCCCTACCCTGTCCGCGTCAACCAGGTTTATGAAAAATG 6240

Qy     6241 GTGCGCCATCACGCCTGAGGCGATGGACAAATCCGGAGCAAATTATGATTCTAAGGTGAT 6300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6241 GTGCGCCATCACGCCTGAGGCGATGGACAAATCCGGAGCAAATTATGATTCTAAGGTGAT 6300

Qy     6301 CAGGTTGCTGGAGCTGTCCTTCCTCGCGGACCGTGAGATGAATACATGGGCATTGCTGAG 6360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6301 CAGGTTGCTGGAGCTGTCCTTCCTCGCGGACCGTGAGATGAATACATGGGCATTGCTGAG 6360

Qy     6361 GGCTAGCACGGCTTTCAAGCTGTACTACCACAAGAGCCCCAAGTTCGTGTGGCAGATGGC 6420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6361 GGCTAGCACGGCTTTCAAGCTGTACTACCACAAGAGCCCCAAGTTCGTGTGGCAGATGGC 6420

Qy     6421 GGGCAGACAGCTCGCGTACATTAAGGCGCAGATGACGAGCAGACCCGGTGAAGGCGCCCC 6480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   6421 GGGCAGACAGCTCGCGTACATTAAGGCGCAGATGACGAGCAGACCCGGTGAAGGCGCCCC 6480

Qy   6481 GGCGTTGATGACCGCGTTCATGTATGCGGGCTTGATGCCGGATAAGAAGTTTACGAAGCA 6540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6481 GGCGTTGATGACCGCGTTCATGTATGCGGGCTTGATGCCGGATAAGAAGTTTACGAAGCA 6540

Qy   6541 GTATGTGGCCAGGCTGGAGGGCGATGGATCGGAGTACCCTGATCCGGAGGTCTATGAAGT 6600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6541 GTATGTGGCCAGGCTGGAGGGCGATGGATCGGAGTACCCTGATCCGGAGGTCTATGAAGT 6600

Qy   6601 GCTGGGCGATGATGATTTTGATGGAATTGGTTTCACAGGGAATGGCGATTATTAGTCGGG 6660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6601 GCTGGGCGATGATGATTTTGATGGAATTGGTTTCACAGGGAATGGCGATTATTAGTCGGG 6660

Qy   6661 TTGCTTGGAGGGGGCCGCTGCATCCTTATTTTTCTGATTTTCCGAGTCTTTGGTGCTGGG 6720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6661 TTGCTTGGAGGGGGCCGCTGCATCCTTATTTTTCTGATTTTCCGAGTCTTTGGTGCTGGG 6720

Qy   6721 AATGGGAATCTTCTGACCATGTTCCCTATGGGTGAGTGCTCATTTTATGGACTTCAGTTT 6780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6721 AATGGGAATCTTCTGACCATGTTCCCTATGGGTGAGTGCTCATTTTATGGACTTCAGTTT 6780

Qy   6781 TGGAGTTTTCTTTCGGCTTTGAGATCTGCCTGAGCAAGCGAGGCGTGGGACATTCGGTTG 6840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6781 TGGAGTTTTCTTTCGGCTTTGAGATCTGCCTGAGCAAGCGAGGCGTGGGACATTCGGTTG 6840

Qy   6841 GTCGCTAGAACATAAAACGAAAATGGGAGATGATATGATTCTTCGTGGCGTGATAATATT 6900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6841 GTCGCTAGAACATAAAACGAAAATGGGAGATGATATGATTCTTCGTGGCGTGATAATATT 6900

Qy   6901 GAGATATTTGGGTGCTGTGCTTTGTATCCATGTCGCAGATATTCGTTCCTACTTGTTGCC 6960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6901 GAGATATTTGGGTGCTGTGCTTTGTATCCATGTCGCAGATATTCGTTCCTACTTGTTGCC 6960

Qy   6961 CGCGAAGCAGATTTGGTCCAACTTTCAAACAATAGTCAGCAACCTTGAGGAAACTATCGA 7020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6961 CGCGAAGCAGATTTGGTCCAACTTTCAAACAATAGTCAGCAACCTTGAGGAAACTATCGA 7020

Qy   7021 TCTCGATACAGACGCTTAGAATAAATAGCGTCCTCGCTCCCTTAGCCTCCGACCTCCATC 7080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7021 TCTCGATACAGACGCTTAGAATAAATAGCGTCCTCGCTCCCTTAGCCTCCGACCTCCATC 7080

Qy   7081 CTAGGTGGAGCACCCGGGATCTCCTAGACTCCTTTCAGACCTGTCTAAGTGAGATATATA 7140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7081 CTAGGTGGAGCACCCGGGATCTCCTAGACTCCTTTCAGACCTGTCTAAGTGAGATATATA 7140

Qy   7141 GTGCAAAGCAGTTGGTGTTACACTTTGGTATTACATCAACAAGACAATTTGAGATAGACT 7200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7141 GTGCAAAGCAGTTGGTGTTACACTTTGGTATTACATCAACAAGACAATTTGAGATAGACT 7200

Qy   7201 CATTCCCTCCGGTTTTCTTCCGATTGCGACTCTATTTGATGATTTTTTTTCCATCTACCC 7260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7201 CATTCCCTCCGGTTTTCTTCCGATTGCGACTCTATTTGATGATTTTTTTTCCATCTACCC 7260

Qy   7261 CTGACAAGCATAACTCTAGCAAATCTTTTTGCTGAATGAATGTAGGCTCTTTGGGATTTC 7320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7261 CTGACAAGCATAACTCTAGCAAATCTTTTTGCTGAATGAATGTAGGCTCTTTGGGATTTC 7320

Qy   7321 GAGGTTTTGACCTTGAAACCTTTGATCTTTTGTTCTTTTGCCTTTTCGTTGAGTTGGAAA 7380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7321 GAGGTTTTGACCTTGAAACCTTTGATCTTTTGTTCTTTTGCCTTTTCGTTGAGTTGGAAA 7380

Qy   7381 GTATATCAACTCTACCTGATCACATCGGAAGCAACAACAACAACAACAACAACAACAACA 7440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7381 GTATATCAACTCTACCTGATCACATCGGAAGCAACAACAACAACAACAACAACAACAACA 7440

Qy   7441 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7441 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7500

Qy   7501 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7501 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7560

Qy   7561 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7561 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7620

Qy   7621 ACAACAACAACAACAACAACAACAACAGCAACAACAACAACAACAACAACAACAACAACA 7680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7621 ACAACAACAACAACAACAACAACAACAGCAACAACAACAACAACAACAACAACAACAACA 7680

Qy   7681 ACAACAACAAGACAACAACAAGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAG 7740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7681 ACAACAACAAGACAACAACAAGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAG 7740

Qy   7741 ACCGAGCGACCTCATAGCTTATGGGAAAATAATACACGTCCAGAGATACATAACAAATAC 7800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7741 ACCGAGCGACCTCATAGCTTATGGGAAAATAATACACGTCCAGAGATACATAACAAATAC 7800

Qy   7801 AGCAACAACAATGAAGTTCAAAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCAC 7860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7801 AGCAACAACAATGAAGTTCAAAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCAC 7860

Qy   7861 CACCACCCACAGCATAAACCTCGAGTGTATGAGTGACGAGTACAACATGAACCGTCGACG 7920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7861 CACCACCCACAGCATAAACCTCGAGTGTATGAGTGACGAGTACAACATGAACCGTCGACG 7920

Qy   7921 ATCCCTATAGTGAGTCGTATTATGCGGCCGCGAATTCGATATCAAGCTTATCGATACCGT 7980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7921 ATCCCTATAGTGAGTCGTATTATGCGGCCGCGAATTCGATATCAAGCTTATCGATACCGT 7980

Qy   7981 CGACCTCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACAATCCCA 8040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7981 CGACCTCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACAATCCCA 8040

Qy   8041 GTCAC 8045
          |||||
Db   8041 GTCAC 8045
```

```
RESULT 2
AAV20700
ID   AAV20700 standard; DNA; 5163 BP.
XX
AC   AAV20700;
XX
DT   17-AUG-1998  (first entry)
XX
DE   Cryptosporidium parvum GP900 antigen open reading frame.
XX
KW   Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW   antibody; prophylaxis; treatment; inhibition; retardation; detection;
KW   diagnosis; human; ds.
XX
OS   Cryptosporidium parvum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..5163
FT                   /*tag=  a
FT                   /product= "GP900 antigen"
FT   misc_feature    524..1270
FT                   /*tag=  b
FT                   /note= "region containing NINC mutations"
XX
PN   WO9806430-A1.
XX
PD   19-FEB-1998.
XX
PF   11-AUG-1997;    97WO-US014104.
XX
PR   14-AUG-1996;    96US-00700651.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Petersen C,  Leech J,  Nelson RC,  Gut J;
```

```
XX
DR   WPI; 1998-159290/14.
DR   P-PSDB; AAW48299.
XX
PT   Anti-Cryptosporidium antibody - used to develop products for detection,
PT   diagnosis, prophylaxis or treatment of Cryptosporidium infections.
XX
PS   Claim 32; Page 60-62; 89pp; English.
XX
CC   The sequence is that encoding the GP900 antigen which may be used in the
CC   production of anti-Cryptosporidium antibodies. These can be used for the
CC   prophylaxis, treatment, inhibition or retardation of a Cryptosporidium
CC   infection in humans or in animals such as calves. They can also be used
CC   for the detection and diagnosis of related infections
XX
SQ   Sequence 5163 BP; 1875 A; 1137 C; 873 G; 1278 T; 0 U; 0 Other;

  Query Match             2.7%;  Score 220.8;  DB 2;  Length 5163;
  Best Local Similarity   67.0%;  Pred. No. 2.8e-36;
  Matches  343;  Conservative    0;  Mismatches  167;  Indels    2;  Gaps    2;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db        668 ACTACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 727

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db        728 ACAACAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACAACCACA 787

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db        788 ACAACTACAACCACAACTACAACCACAACAACTACAACCACAACAACAACAACCACAACA 847

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db        848 ACTACAACCACAACTACAACCACAACAACTACAACCACAACAACCACAACCACAACCACA 907

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db        908 ACTACCAAGAAACCAACAACAACAACTACAACAACAACAACAACAACAACA-ACTACTAC 966

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db        967 TACAACCACCACAACAACAACAACAACAACAACTACAACTACCAAGAAACCAACAACTAC 1026

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db       1027 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAA-CAACTACTA 1085

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881
Db       1086 CTACTACTACAACAACAACAACAACAACAACAACAACAACAACTACCACGAAACCAACAA 1145

Qy       7882 CGAGTGTATGAGTGACGAGTACAACATGAACC 7913
Db       1146 CAACAACAACAACTACTACTACTACTACAACC 1177


RESULT 3
AAA61849
ID   AAA61849 standard; DNA; 5163 BP.
XX
AC   AAA61849;
XX
DT   28-OCT-2000  (first entry)
XX
DE   ORF encoding a portion of Cryptosporidium parvum NINC isolate GP900.
XX
KW   GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis;
KW   competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW   merozoite; diarrhoea; protozoacide; open reading frame; ORF; ds.
XX
OS   Cryptosporidium parvum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1669. .7182
FT                   /*tag=  a
FT                   /partial
FT                   /product= "Cryptosporidium parvum NINC isolate GP900"
FT                   /note= "No start or stop codons given in the
FT                   specification"
XX
PN   US6071518-A.
XX
PD   06-JUN-2000.
XX
PF   12-SEP-1997;   97US-00928361.
XX
PR   29-MAY-1992;   92US-00891301.
PR   01-JUN-1993;   93US-00071880.
PR   03-APR-1995;   95US-00415751.
PR   14-AUG-1996;   96US-00700651.
PR   13-SEP-1996;   96US-0026062P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Petersen C;
XX
DR   WPI; 2000-422065/36.
DR   P-PSDB; AAB11727.
XX
PT   New GP900 protein fragments and fusion proteins of Cryptosporidium
PT   parvum, useful for detecting the presence of the parasite, and diagnosing
PT   or treating Cryptosporidium infections by competitive inhibition of the
PT   function of GP900.
XX
PS   Claim 16; Col 47-52; 59pp; English.
XX
CC   The invention relates to the GP900 glycoprotein of the protozoan
CC   Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC   proteins comprising GP900 fragments. The invention also relates to the
CC   administration of GP900 or fragments thereof to a host to elicit anti-
CC   GP900 antibody production, and to a method of cryptosporidiosis treatment
CC   or prophylaxis comprising administration of anti-GP900 antibodies to an
CC   individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC   competitively inhibit sporozoite or merozoite attachment or invasion, and
CC   are also useful for the generation of anti-GP900 antibodies. The
CC   antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC   additionally inhibit the binding of GP900 ligands to GP900. GP900
CC   proteins, fragments and antibodies may therefore be used to treat or
CC   prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC   cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC   immunocompromised persons. Cryptosporidiosis can be contracted from
CC   contaminated municipal water supplies (e.g., public swimming pools). It
CC   is also a cause of disease in animals, resulting in financial losses in
CC   agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC   used for the diagnosis of Cryptosporidium parvum infections, and for the
CC   detection of the parasite in the environment. The present sequence
CC   represents the open reading frame (ORF) encoding a portion of the GP900
CC   protein of the NINC isolate of Cryptosporidium parvum
XX
SQ   Sequence 5163 BP; 1873 A; 1138 C; 875 G; 1277 T; 0 U; 0 Other;

  Query Match             2.7%;  Score 220.8;  DB 3;  Length 5163;
  Best Local Similarity   67.0%;  Pred. No. 2.8e-36;
  Matches  343;  Conservative    0;  Mismatches  167;  Indels    2;  Gaps    2;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db        667 ACTACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 726

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db        727 ACAACAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACAACCACA 786

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
```

```
Db        787 ACAACTACAACCACAACTACAACCACAACAACTACAACCACAACAACAACAACCACAACA 846

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641

Db        847 ACTACAACCACAACTACAACCACAACAACTACAACCACAACAACCACAACCACAACCACA 906

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701

Db        907 ACTACCAAGAAACCAACAACAACAACTACAACAACAACAACAACAACAACA-ACTACTAC 965

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761

Db        966 TACAACCACCACAACAACAACAACAACAACAACTACAACTACCAAGAAACCAACAACTAC 1025

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821

Db       1026 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAA-CAACTACTA 1084

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881

Db       1085 CTACTACTACAACAACAACAACAACAACAACAACAACAACAACTACCACGAAACCAACAA 1144

Qy       7882 CGAGTGTATGAGTGACGAGTACAACATGAACC 7913

Db       1145 CAACAACAACAACTACTACTACTACTACAACC 1176
```

RESULT 4
ABT04778
ID   ABT04778 standard; DNA; 5163 BP.
XX
AC   ABT04778;
XX
DT   27-SEP-2002 (first entry)
XX
DE   C parvum GP900 gene fragment SEQ ID NO: 4.
XX
KW   Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;
KW   gene; ds.
XX
OS   Cryptosporidium parvum.
XX
PN   WO200194631-A1.
XX
PD   13-DEC-2001.
XX
PF   14-MAY-2001; 2001WO-US015624.
XX
PR   06-JUN-2000; 2000US-00588995.
XX
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Petersen C, Barnes DA, Nelson RG, Gut J;
XX
DR   WPI; 2002-566447/60.
XX
PT   Detecting Cryptosporidium in biological and environmental samples and
PT   diagnosis of cryptosporidiosis involves, contacting the sample with
PT   Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA.
XX
PS   Disclosure; Page 103-104; 157pp; English.
XX
CC   The present invention relates to a method of detecting Cryptosporidium in
CC   biological and environmental samples, and of diagnosing
CC   cryptosporidiosis. This involves obtaining a sample and contacting it
CC   with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or
CC   RNA, or its variant, mutant or fragment. The method is also useful for
CC   detecting and identifying individual Cryptosporidium isolates based on
CC   the genetic characteristics, and for diagnosis of prior or concurrent
CC   Cryptosporidium infection. The present sequence is a C. parvum coding
CC   sequence used in the exemplification of the invention
XX
SQ   Sequence 5163 BP; 1873 A; 1138 C; 875 G; 1277 T; 0 U; 0 Other;

```
  Query Match             2.7%;  Score 220.8;  DB 6;  Length 5163;
  Best Local Similarity   67.0%;  Pred. No. 2.8e-36;
  Matches  343;  Conservative    0;  Mismatches  167;  Indels    2;  Gaps    2;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461

Db        667 ACTACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 726

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521

Db        727 ACAACAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACAACCACA 786

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581

Db        787 ACAACTACAACCACAACTACAACCACAACAACTACAACCACAACAACAACAACCACAACA 846

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641

Db        847 ACTACAACCACAACTACAACCACAACAACTACAACCACAACAACCACAACCACAACCACA 906

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701

Db        907 ACTACCAAGAAACCAACAACAACAACTACAACAACAACAACAACAACAACA-ACTACTAC 965

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761

Db        966 TACAACCACCACAACAACAACAACAACAACAACTACAACTACCAAGAAACCAACAACTAC 1025

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821

Db       1026 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAA-CAACTACTA 1084

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881

Db       1085 CTACTACTACAACAACAACAACAACAACAACAACAACAACAACTACCACGAAACCAACAA 1144

Qy       7882 CGAGTGTATGAGTGACGAGTACAACATGAACC 7913

Db       1145 CAACAACAACAACTACTACTACTACTACAACC 1176
```

RESULT 5
AAA61848
ID   AAA61848 standard; DNA; 5318 BP.
XX
AC   AAA61848;
XX
DT   28-OCT-2000 (first entry)
XX
DE   DNA encoding a portion of Cryptosporidium parvum NINC isolate GP900.
XX
KW   GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis;
KW   competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW   merozoite; diarrhoea; protozoacide; ds.
XX
OS   Cryptosporidium parvum.
XX
FH   Key   Location/Qualifiers
FT   CDS   1669. .7182
FT   /*tag= a
FT   /partial
FT   /product= "Cryptosporidium parvum NINC isolate GP900"
FT   /note= "No start codon given in the specification"
FT   3'UTR   5167. .5318
FT   /*tag= b
XX
PN   US6071518-A.
XX
PD   06-JUN-2000.
XX
PF   12-SEP-1997;   97US-00928361.

```
XX
PR   29-MAY-1992;   92US-00891301.
PR   01-JUN-1993;   93US-00071880.
PR   03-APR-1995;   95US-00415751.
PR   14-AUG-1996;   96US-00700651.
PR   13-SEP-1996;   96US-0026062P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Petersen C;
XX
DR   WPI; 2000-422065/36.
DR   P-PSDB; AAB11727.
XX
PT   New GP900 protein fragments and fusion proteins of Cryptosporidium
PT   parvum, useful for detecting the presence of the parasite, and diagnosing
PT   or treating Cryptosporidium infections by competitive inhibition of the
PT   function of GP900.
XX
PS   Claim 16; Col 41-48; 59pp; English.
XX
CC   The invention relates to the GP900 glycoprotein of the protozoan
CC   Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC   proteins comprising GP900 fragments. The invention also relates to the
CC   administration of GP900 or fragments thereof to a host to elicit anti-
CC   GP900 antibody production, and to a method of cryptosporidiosis treatment
CC   or prophylaxis comprising administration of anti-GP900 antibodies to an
CC   individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC   competitively inhibit sporozoite or merozoite attachment or invasion, and
CC   are also useful for the generation of anti-GP900 antibodies. The
CC   antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC   additionally inhibit the binding of GP900 ligands to GP900. GP900
CC   proteins, fragments and antibodies may therefore be used to treat or
CC   prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC   cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC   immunocompromised persons. Cryptosporidiosis can be contracted from
CC   contaminated municipal water supplies (e.g., public swimming pools). It
CC   is also a cause of disease in animals, resulting in financial losses in
CC   agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC   used for the diagnosis of Cryptosporidium parvum infections, and for the
CC   detection of the parasite in the environment. The present sequence
CC   represents genomic DNA encoding a portion of the GP900 protein of the
CC   NINC isolate of Cryptosporidium parvum
XX
SQ   Sequence 5318 BP; 1939 A; 1158 C; 890 G; 1331 T; 0 U; 0 Other;

  Query Match             2.7%;  Score 220.8;  DB 3;  Length 5318;
  Best Local Similarity   67.0%;  Pred. No. 2.9e-36;
  Matches  343;  Conservative    0;  Mismatches  167;  Indels    2;  Gaps    2;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db        667 ACTACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 726

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db        727 ACAACAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACAACCACA 786

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db        787 ACAACTACAACCACAACTACAACCACAACAACTACAACCACAACAACAACAACCACAACA 846

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db        847 ACTACAACCACAACTACAACCACAACAACTACAACCACAACAACCACAACCACAACCACA 906

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db        907 ACTACCAAGAAACCAACAACAACAACTACAACAACAACAACAACAACA-ACTACTAC 965

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db        966 TACAACCACCACAACAACAACAACAACAACAACTACAACTACCAAGAAACCAACAACTAC 1025

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db       1026 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAA-CAACTACTA 1084

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881
Db       1085 CTACTACTACAACAACAACAACAACAACAACAACAACAACAACTACCACGAAACCAACAA 1144

Qy       7882 CGAGTGTATGAGTGACGAGTACAACATGAACC 7913
Db       1145 CAACAACAACAACTACTACTACTACTACAACC 1176


RESULT 6
ABT04777
ID   ABT04777 standard; DNA; 5318 BP.
XX
AC   ABT04777;
XX
DT   27-SEP-2002  (first entry)
XX
DE   C parvum GP900 gene fragment SEQ ID NO: 3.
XX
KW   Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;
KW   gene; ds.
XX
OS   Cryptosporidium parvum.
XX
PN   WO200194631-A1.
XX
PD   13-DEC-2001.
XX
PF   14-MAY-2001; 2001WO-US015624.
XX
PR   06-JUN-2000; 2000US-00588995.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Petersen C,  Barnes DA,  Nelson RG,  Gut J;
XX
DR   WPI; 2002-566447/60.
XX
PT   Detecting Cryptosporidium in biological and environmental samples and
PT   diagnosis of cryptosporidiosis involves, contacting the sample with
PT   Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA.
XX
PS   Disclosure; Page 101-103; 157pp; English.
XX
CC   The present invention relates to a method of detecting Cryptosporidium in
CC   biological and environmental samples, and of diagnosing
CC   cryptosporidiosis. This involves obtaining a sample and contacting it
CC   with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or
CC   RNA, or its variant, mutant or fragment. The method is also useful for
CC   detecting and identifying individual Cryptosporidium isolates based on
CC   the genetic characteristics, and for diagnosis of prior or concurrent
CC   Cryptosporidium infection. The present sequence is a C. parvum coding
CC   sequence used in the exemplification of the invention
XX
SQ   Sequence 5318 BP; 1939 A; 1158 C; 890 G; 1331 T; 0 U; 0 Other;

  Query Match             2.7%;  Score 220.8;  DB 6;  Length 5318;
  Best Local Similarity   67.0%;  Pred. No. 2.9e-36;
  Matches  343;  Conservative    0;  Mismatches  167;  Indels    2;  Gaps    2;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db        667 ACTACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 726

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db        727 ACAACAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACAACCACA 786
```

```
Qy      7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db       787 ACAACTACAACCACAACTACAACCACAACAACTACAACCACAACAACAACAACCACAACA 846

Qy      7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db       847 ACTACAACCACAACTACAACCACAACAACTACAACCACAACAACCACAACCACAACCACA 906

Qy      7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db       907 ACTACCAAGAAACCAACAACAACAACTACAACAACAACAACAACAACAACA-ACTACTAC 965

Qy      7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db       966 TACAACCACCACAACAACAACAACAACAACAACTACAACTACCAAGAAACCAACAACTAC 1025

Qy      7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db      1026 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAA-CAACTACTA 1084

Qy      7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881
Db      1085 CTACTACTACAACAACAACAACAACAACAACAACAACAACAACTACCACGAAACCAACAA 1144

Qy      7882 CGAGTGTATGAGTGACGAGTACAACATGAACC 7913
Db      1145 CAACAACAACAACTACTACTACTACTACAACC 1176
```

```
RESULT 7
AAV20701
ID   AAV20701 standard; DNA; 5318 BP.
XX
AC   AAV20701;
XX
DT   17-AUG-1998  (first entry)
XX
DE   Cryptosporidium parvum GP900 antigen open reading frame and 3' region.
XX
KW   Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW   antibody; prophylaxis; treatment; inhibition; retardation; detection;
KW   diagnosis; human; 3' region; ds.
XX
OS   Cryptosporidium parvum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..5166
FT                   /*tag=  a
FT                   /product= "GP900 antigen"
FT   misc_feature    524..1270
FT                   /*tag=  b
FT                   /note= "region containing NINC mutations"
XX
PN   WO9806430-A1.
XX
PD   19-FEB-1998.
XX
PF   11-AUG-1997;   97WO-US014104.
XX
PR   14-AUG-1996;   96US-00700651.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Petersen C,  Leech J,  Nelson RC,  Gut J;
XX
DR   WPI; 1998-159290/14.
DR   P-PSDB; AAW48299.
XX
PT   Anti-Cryptosporidium antibody - used to develop products for detection,
PT   diagnosis, prophylaxis or treatment of Cryptosporidium infections.
XX
PS   Claim 32; Page 62-63; 89pp; English.
XX
CC   The sequence is that encoding the GP900 antigen which may be used in the
CC   production of anti-Cryptosporidium antibodies. These can be used for the
CC   prophylaxis, treatment, inhibition or retardation of a Cryptosporidium
CC   infection in humans or in animals such as calves. They can also be used
CC   for the detection and diagnosis of related infections
XX
SQ   Sequence 5318 BP; 1944 A; 1153 C; 888 G; 1333 T; 0 U; 0 Other;

  Query Match             2.7%;  Score 219.2;  DB 2;  Length 5318;
  Best Local Similarity   66.8%;  Pred. No. 6.2e-36;
  Matches 342;  Conservative    0;  Mismatches  168;  Indels    2;  Gaps    2;

Qy      7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db       668 ACTACGACAACAACAACAAAAACAACAACAACAACAACAACAACAACAACAACAACAACA 727

Qy      7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db       728 ACAACAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACAACCACA 787

Qy      7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db       788 ACAACTACAACCACAACTACAACCACAACAACTACAACCACAACAACAACAACCACAACA 847

Qy      7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db       848 ACTACAACCACAACTACAACCACAACAACTACAACCACAACAACCACAACCACAACCACA 907

Qy      7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db       908 ACTACCAAGAAACCAACAACAACAACTACAACAACAACAACAACAACAACA-ACTACTAC 966

Qy      7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db       967 TACAACCACCACAACAACAACAACAACAACAACTACAACTACCAAGAAACCAACAACTAC 1026

Qy      7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db      1027 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAA-CAACTACTA 1085

Qy      7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881
Db      1086 CTACTACTACAACAACAACAACAACAACAACAACAACAACAACTACCACGAAACCAACAA 1145

Qy      7882 CGAGTGTATGAGTGACGAGTACAACATGAACC 7913
Db      1146 CAACAACAACAACTACTACTACTACTACAACC 1177

RESULT 8
AAD44410
ID   AAD44410 standard; DNA; 1236 BP.
XX
AC   AAD44410;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Human huntington (htQ103) protein encoding DNA.
XX
KW   Human; protein misfolding; Alzheimer's disease; AD; Parkinson's disease;
KW   PD; Familial amyloid polyneuropathy; tauopathy; frontotemporal dementia;
KW   Pick disease; lobar atrophy; trinucleotide disease; fragile-X syndrome;
KW   Huntington's disease; spinocerebellar ataxia; SCA; myotonic dystrophy;
KW   dentatorubral pallidoluysian atrophy; DRPLA; Creutzfeldt-Jacob disease;
KW   CJD; prion disease; Gerstmann-Straussler-Scheinker disease; GSS; FFI;
KW   fatal familia insomnia; mad cow disease; scrapie; kuru; anticonvulsant;
KW   nootropic; neuroprotective; cerebroprotective; htQ103 protein; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..513
FT                   /*tag=  a
```

```
FT                   /product= "Human htQ103 protein"
FT                   /note= "No stop codon"
FT                   /partial
XX
PN   WO200265136-A2.
XX
PD   22-AUG-2002.
XX
PF   15-FEB-2002; 2002WO-US004632.
XX
PR   15-FEB-2001; 2001US-0269157P.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Lindquist S,  Krobitsch S,  Outeiro T;
XX
DR   WPI; 2002-667026/71.
DR   P-PSDB; AAE26650.
XX
PT   Screening for therapeutic agents for protein misfolding disease, by
PT   contacting a yeast cell with compound, that expresses misfolded disease
PT   protein, and with a toxicity inducing agent, and evaluating cell for
PT   viability.
XX
PS   Disclosure; Page 87-88; 93pp; English.
XX
CC   The present invention relates to novel screening methods for identifying
CC   therapeutic agents for diseases associated with protein misfolding. The
CC   method involves contacting a yeast cell with a candidate compound, where
CC   the yeast cell expresses a polypeptide comprising a misfolded disease
CC   protein, contacting the yeast cell with a toxicity inducing agent and
CC   evaluating the yeast cell for viability, where the viability indicates
CC   the candidate compound is a candidate therapeutic agent. The method is
CC   useful to screen for therapeutic agents for diseases associated with
CC   protein misfolding such as Alzheimer's disease (AD), Parkinson's disease
CC   (PD), Familial amyloid polyneuropathy, tauopathies (e.g. Pick disease,
CC   lobar atrophy, frontotemporal dementia) or trinucleotide diseases (e.g.
CC   Huntington's disease, spinocerebellar ataxia (SCA), fragile-X syndrome,
CC   myotonic dystrophy, dentatorubral pallidoluysian atrophy (DRPLA) and
CC   prion diseases (e.g. Creutzfeldt-Jacob disease (CJD), fatal familia
CC   insomnia (FFI), Gerstmann-Straussler-Scheinker disease (GSS), mad cow
CC   disease, scrapie and kuru). The method is useful for treating a patient
CC   with Huntington's disease or Parkinson's disease. The present sequence is
CC   a DNA encoding human huntington (htQ103) protein. This sequence is used
CC   to illustrate the method of the invention
XX
SQ   Sequence 1236 BP; 368 A; 441 C; 288 G; 139 T; 0 U; 0 Other;

  Query Match             2.6%;  Score 211.4;  DB 6;  Length 1236;
  Best Local Similarity   81.4%;  Pred. No. 1.9e-34;
  Matches  245;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
              ||| |  |  | || ||||||| || ||||| |||||||| || ||||| |||||||| ||
Db         54 ACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCA 113

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
                ||||| ||||||| || ||||| |||||||| || ||||| |||||||| || |||||
Db        114 GCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACA 173

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
                ||||||| || ||||| |||||||| || ||||| |||||||| || ||||| |||||
Db        174 GCAACAACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACAGCAACA 233

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
              ||| || ||||| |||||||| || ||||| |||||||| || ||||| |||||||| ||
Db        234 ACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCA 293

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
                ||||||||||||| || ||||| |||||||| || ||||| |||||| |  |  |  |
Db        294 GCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACA 353

Qy       7702 G 7702
              |
Db        354 G 354


RESULT 9
AAA29550/c
ID   AAA29550 standard; DNA; 567 BP.
XX
AC   AAA29550;
XX
DT   09-AUG-2000  (first entry)
XX
DE   HIV codon altered env nucleotide sequence #2.
XX
KW   Erythropoietin; EPO; G-CSF; granulocyte colony stimulating factor;
KW   wobble; codon altered gene; shuffling; modification; vaccine; insulin;
KW   peptide hormone; growth factor; cytokine; interferon; interleukin;
KW   leukaemia inhibitory factor; oncostatin M; transcription activator;
KW   expression activator; infectious organism; ds.
XX
OS   Human immunodeficiency virus 1.
OS   Synthetic.
XX
PN   WO200018906-A2.
XX
PD   06-APR-2000.
XX
PF   28-SEP-1999;   99WO-US022588.
XX
PR   29-SEP-1998;   98US-0102362P.
PR   29-JAN-1999;   99US-0117729P.
PR   05-FEB-1999;   99US-0118813P.
PR   24-JUN-1999;   99US-0141049P.
XX
PA   (MAXY-) MAXYGEN INC.
XX
PI   Patten PA,  Liu L,  Stemmer WPC;
XX
DR   WPI; 2000-303449/26.
XX
PT   Novel methods for recombining codon-altered libraries of nucleic acids
PT   used to produce new proteins and new vectors with reduced rates of
PT   reversion to wild type.
XX
PS   Example; Fig 18A; 92pp; English.
XX
CC   A method has been developed of making codon altered nucleic acids (NAs),
CC   comprising providing a NA sequence (NA1) which encodes a polypeptide
CC   (P1), providing codon altered NA sequences, each encoding P1 or a
CC   modified form of it, and recombining the codon altered NA sequences to
CC   produce a target codon altered NA which encodes a second protein. The
CC   method of the invention can be used for recombining codon-altered
CC   libraries of nucleic acids to produce new proteins, which have
CC   improvements in a desirable characteristic. Target nucleic acids include
CC   those coding for therapeutic proteins such as erythropoietin (EPO),
CC   insulin, peptide hormones, growth factors, cytokines, interferons,
CC   interleukins, leukaemia inhibitory factor, and oncostatin M, as well as
CC   transcription and expression activators and proteins from infectious
CC   organisms for use as vaccines. The method can also be used to produce
CC   attenuated viruses which have reduced rates of reversion to wild type.
CC   The present sequence represents an HIV codon altered env nucleotide
CC   sequence, which is used in an example from the present invention
XX
SQ   Sequence 567 BP; 32 A; 129 C; 222 G; 184 T; 0 U; 0 Other;

  Query Match             2.5%;  Score 205;  DB 3;  Length 567;
  Best Local Similarity   81.2%;  Pred. No. 3.5e-33;
  Matches  238;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy       7398 GATCACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7457
                || || ||     | || || || || || || || ||| || || || || || || |
Db        440 GACGACGACGACGACGACGACGACGACGACGACGACGACGACAGCAGCAGCAGCAGCAGC 381
```

```
Qy       7458 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7517
Db        380 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 321

Qy       7518 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7577
Db        320 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAACAACAACAACAACAACAACAACAAC 261

Qy       7578 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7637
Db        260 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 201

Qy       7638 AACAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7690
Db        200 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 148
```

RESULT 10
ABL25662/c
ID   ABL25662 standard; DNA; 2336 BP.
XX
AC   ABL25662;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 28459.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 28459; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2336 BP; 554 A; 411 C; 550 G; 821 T; 0 U; 0 Other;

Query Match 2.5%; Score 203.8; DB 4; Length 2336;
Best Local Similarity 74.4%; Pred. No. 8.9e-33;
Matches 270; Conservative 0; Mismatches 92; Indels 1; Gaps 1;

```
Qy       7393 TACCTGATCACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACA 7452
Db       1991 TATCCGATAACAACGTCTCCCAGAATAACAACATCTTCAACTACAACAACACCAACTACA 1932

Qy       7453 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7512
Db       1931 ACTAAAACAACACCAACTACAACAACACCAACTACAACAACACAAACTACAAGAACATCA 1872

Qy       7513 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7572
Db       1871 ACTACAACAACTCCAACTACAACAACAGCAACTACAACACCAACGACAACAACAACAAAA 1812

Qy       7573 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7632
Db       1811 ACTACAACAACTCCAACTACAACTCCAACTTCAACACCAACTACAACAACTACAACAACA 1752

Qy       7633 ACAACAACAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAAC-AAG 7691
Db       1751 ACAACACAAACTACAGGAACATCTACTACAACAACTCCAACTACAACAACACCAACTACA 1692

Qy       7692 ACAACAACAAGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACC 7751
Db       1691 ACAACTCCAACTATAACAACACCAACAACTACAACACCAACTACAACAACAACAACTACA 1632

Qy       7752 TCA 7754
Db       1631 ACA 1629
```

RESULT 11
AAA61847
ID   AAA61847 standard; DNA; 5511 BP.
XX
AC   AAA61847;
XX
DT   28-OCT-2000 (first entry)
XX
DE   Cryptosporidium parvum Iowa isolate GP900 ORF.
XX
KW   GP900; Iowa isolate; glycoprotein; antibody; cryptosporidiosis;
KW   competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW   merozoite; diarrhoea; protozoacide; open reading frame; ORF; ds.
XX
OS   Cryptosporidium parvum.
XX
FH   Key   Location/Qualifiers
FT   CDS   1. .5511
FT   /*tag= a
FT   /product= "Cryptosporidium parvum Iowa isolate GP900"
FT   /note= "No stop codon given"
XX
PN   US6071518-A.
XX
PD   06-JUN-2000.
XX
PF   12-SEP-1997;   97US-00928361.
XX
PR   29-MAY-1992;   92US-00891301.
PR   01-JUN-1993;   93US-00071880.
PR   03-APR-1995;   95US-00415751.
PR   14-AUG-1996;   96US-00700651.
PR   13-SEP-1996;   96US-0026062P.
XX
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Petersen C;
XX
DR   WPI; 2000-422065/36.
DR   P-PSDB; AAB11726.
XX
PT   New GP900 protein fragments and fusion proteins of Cryptosporidium
PT   parvum, useful for detecting the presence of the parasite, and diagnosing
PT   or treating Cryptosporidium infections by competitive inhibition of the
PT   function of GP900.
XX
PS   Claim 16; Col 37-42; 59pp; English.

```
XX
CC   The invention relates to the GP900 glycoprotein of the protozoan
CC   Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC   proteins comprising GP900 fragments. The invention also relates to the
CC   administration of GP900 or fragments thereof to a host to elicit anti-
CC   GP900 antibody production, and to a method of cryptosporidiosis treatment
CC   or prophylaxis comprising administration of anti-GP900 antibodies to an
CC   individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC   competitively inhibit sporozoite or merozoite attachment or invasion, and
CC   are also useful for the generation of anti-GP900 antibodies. The
CC   antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC   additionally inhibit the binding of GP900 ligands to GP900. GP900
CC   proteins, fragments and antibodies may therefore be used to treat or
CC   prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC   cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC   immunocompromised persons. Cryptosporidiosis can be contracted from
CC   contaminated municipal water supplies (e.g., public swimming pools). It
CC   is also a cause of disease in animals, resulting in financial losses in
CC   agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC   used for the diagnosis of Cryptosporidium parvum infections, and for the
CC   detection of the parasite in the environment. The present sequence
CC   represents the open erading frame (ORF) encoding the GP900 protein of the
CC   Iowa isolate of Cryptosporidium parvum
XX
SQ   Sequence 5511 BP; 1964 A; 1204 C; 944 G; 1399 T; 0 U; 0 Other;

  Query Match             2.5%;  Score 200.4;  DB 3;  Length 5511;
  Best Local Similarity   65.5%;  Pred. No. 5.7e-32;
  Matches  308;  Conservative    0;  Mismatches  161;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db       1018 ACTACGACAACAACAACTACTACTACGACAACAACAACAACAACTACTACTACTACTACA 1077

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db       1078 ACCACAACAACTACAACCACGACAACTACAACCACAACCACAACTACCAAGAAACCAACA 1137

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db       1138 ACAACAACAACAACAACAACTACTACTACTACAACAACAACAACAACAACTACTACTACT 1197

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db       1198 ACTACTACTACTACTACTACTACTACTACTACTACCACAACAACCACAACCACAACCACA 1257

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db       1258 ACTACCAAGAAACCAACAACAACAACAACAACAACAACAACAACTACTA-CTACAACCAC 1316

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db       1317 GACAACAACAACCACGACAACCACAACCACAACTACAACTACCAAGAAACCAACAACTAC 1376

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db       1377 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAACAACTACTAC 1436

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACA 7871
Db       1437 TACTACTACAACCACAACCACAACCACAACCGCAACCACAACTACCAAGA 1486

RESULT 12
ABT04776
ID   ABT04776 standard; DNA; 5511 BP.
XX
AC   ABT04776;
XX
DT   27-SEP-2002  (first entry)
XX
DE   C parvum GP900 gene fragment SEQ ID NO: 2.
XX
KW   Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;
KW   gene; ds.
XX
OS   Cryptosporidium parvum.
XX
PN   WO200194631-A1.
XX
PD   13-DEC-2001.
XX
PF   14-MAY-2001; 2001WO-US015624.
XX
PR   06-JUN-2000; 2000US-00588995.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Petersen C,  Barnes DA,  Nelson RG,  Gut J;
XX
DR   WPI; 2002-566447/60.
XX
PT   Detecting Cryptosporidium in biological and environmental samples and
PT   diagnosis of cryptosporidiosis involves, contacting the sample with
PT   Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA.
XX
PS   Disclosure; Page 99-101; 157pp; English.
XX
CC   The present invention relates to a method of detecting Cryptosporidium in
CC   biological and environmental samples, and of diagnosing
CC   cryptosporidiosis. This involves obtaining a sample and contacting it
CC   with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or
CC   RNA, or its variant, mutant or fragment. The method is also useful for
CC   detecting and identifying individual Cryptosporidium isolates based on
CC   the genetic characteristics, and for diagnosis of prior or concurrent
CC   Cryptosporidium infection. The present sequence is a C. parvum coding
CC   sequence used in the exemplification of the invention
XX
SQ   Sequence 5511 BP; 1964 A; 1204 C; 944 G; 1399 T; 0 U; 0 Other;

  Query Match             2.5%;  Score 200.4;  DB 6;  Length 5511;
  Best Local Similarity   65.5%;  Pred. No. 5.7e-32;
  Matches  308;  Conservative    0;  Mismatches  161;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db       1018 ACTACGACAACAACAACTACTACTACGACAACAACAACAACAACTACTACTACTACTACA 1077

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db       1078 ACCACAACAACTACAACCACGACAACTACAACCACAACCACAACTACCAAGAAACCAACA 1137

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db       1138 ACAACAACAACAACAACAACTACTACTACTACAACAACAACAACAACAACTACTACTACT 1197

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db       1198 ACTACTACTACTACTACTACTACTACTACTACTACCACAACAACCACAACCACAACCACA 1257

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db       1258 ACTACCAAGAAACCAACAACAACAACAACAACAACAACAACAACTACTA-CTACAACCAC 1316

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db       1317 GACAACAACAACCACGACAACCACAACCACAACTACAACTACCAAGAAACCAACAACTAC 1376

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db       1377 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAACAACTACTAC 1436

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACA 7871
Db       1437 TACTACTACAACCACAACCACAACCACAACCGCAACCACAACTACCAAGA 1486
```

```
RESULT 13
AAA61846
ID   AAA61846 standard; DNA; 7334 BP.
XX
AC   AAA61846;
XX
DT   28-OCT-2000  (first entry)
XX
DE   Cryptosporidium parvum Iowa isolate GP900 DNA.
XX
KW   GP900; Iowa isolate; glycoprotein; antibody; cryptosporidiosis;
KW   competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW   merozoite; diarrhoea; protozoacide; ds.
XX
OS   Cryptosporidium parvum.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1. .1668
FT                   /*tag=  a
FT   CDS             1669. .7182
FT                   /*tag=  b
FT                   /product= "Cryptosporidium parvum Iowa isolate GP900"
FT                   /transl_except= (pos:7021. .7023, aa:Asp)
FT   3'UTR           7183. .7334
FT                   /*tag=  c
XX
PN   US6071518-A.
XX
PD   06-JUN-2000.
XX
PF   12-SEP-1997;   97US-00928361.
XX
PR   29-MAY-1992;   92US-00891301.
PR   01-JUN-1993;   93US-00071880.
PR   03-APR-1995;   95US-00415751.
PR   14-AUG-1996;   96US-00700651.
PR   13-SEP-1996;   96US-0026062P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Petersen C;
XX
DR   WPI; 2000-422065/36.
DR   P-PSDB; AAB11726.
XX
PT   New GP900 protein fragments and fusion proteins of Cryptosporidium
PT   parvum, useful for detecting the presence of the parasite, and diagnosing
PT   or treating Cryptosporidium infections by competitive inhibition of the
PT   function of GP900.
XX
PS   Claim 16; Col 31-38; 59pp; English.
XX
CC   The invention relates to the GP900 glycoprotein of the protozoan
CC   Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC   proteins comprising GP900 fragments. The invention also relates to the
CC   administration of GP900 or fragments thereof to a host to elicit anti-
CC   GP900 antibody production, and to a method of cryptosporidiosis treatment
CC   or prophylaxis comprising administration of anti-GP900 antibodies to an
CC   individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC   competitively inhibit sporozoite or merozoite attachment or invasion, and
CC   are also useful for the generation of anti-GP900 antibodies. The
CC   antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC   additionally inhibit the binding of GP900 ligands to GP900. GP900
CC   proteins, fragments and antibodies may therefore be used to treat or
CC   prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC   cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC   immunocompromised persons. Cryptosporidiosis can be contracted from
CC   contaminated municipal water supplies (e.g., public swimming pools). It
CC   is also a cause of disease in animals, resulting in financial losses in
CC   agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC   used for the diagnosis of Cryptosporidium parvum infections, and for the
CC   detection of the parasite in the environment. The present sequence
CC   represents genomic DNA encoding the GP900 protein of the Iowa isolate of
CC   Cryptosporidium parvum
XX
SQ   Sequence 7334 BP; 2629 A; 1536 C; 1261 G; 1908 T; 0 U; 0 Other;

  Query Match             2.5%;  Score 200.4;  DB 3;  Length 7334;
  Best Local Similarity   65.5%;  Pred. No. 6.1e-32;
  Matches  308;  Conservative    0;  Mismatches  161;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461

Db       2686 ACTACGACAACAACAACTACTACTACGACAACAACAACAACAACTACTACTACTACTACA 2745

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521

Db       2746 ACCACAACAACTACAACCACGACAACTACAACCACAACCACAACTACCAAGAAACCAACA 2805

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581

Db       2806 ACAACAACAACAACAACAACTACTACTACTACAACAACAACAACAACAACTACTACTACT 2865

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641

Db       2866 ACTACTACTACTACTACTACTACTACTACTACTACCACAACAACCACAACCACAACCACA 2925

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701

Db       2926 ACTACCAAGAAACCAACAACAACAACAACAACAACAACAACAACTACTA-CTACAACCAC 2984

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761

Db       2985 GACAACAACAACCACGACAACCACAACCACAACTACAACTACCAAGAAACCAACAACTAC 3044

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821

Db       3045 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAACAACTACTAC 3104

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACA 7871

Db       3105 TACTACTACAACCACAACCACAACCACAACCGCAACCACAACTACCAAGA 3154


RESULT 14
ABT04775
ID   ABT04775 standard; DNA; 7334 BP.
XX
AC   ABT04775;
XX
DT   27-SEP-2002  (first entry)
XX
DE   C parvum GP900 gene fragment SEQ ID NO: 1.
XX
KW   Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;
KW   gene; ds.
XX
OS   Cryptosporidium parvum.
XX
PN   WO200194631-A1.
XX
PD   13-DEC-2001.
XX
PF   14-MAY-2001; 2001WO-US015624.
XX
PR   06-JUN-2000; 2000US-00588995.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Petersen C,  Barnes DA,  Nelson RG,  Gut J;
XX
DR   WPI; 2002-566447/60.
XX
PT   Detecting Cryptosporidium in biological and environmental samples and
PT   diagnosis of cryptosporidiosis involves, contacting the sample with
```

```
PT   Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA.
XX
PS   Disclosure; Page 97-99; 157pp; English.
XX
CC   The present invention relates to a method of detecting Cryptosporidium in
CC   biological and environmental samples, and of diagnosing
CC   cryptosporidiosis. This involves obtaining a sample and contacting it
CC   with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or
CC   RNA, or its variant, mutant or fragment. The method is also useful for
CC   detecting and identifying individual Cryptosporidium isolates based on
CC   the genetic characteristics, and for diagnosis of prior or concurrent
CC   Cryptosporidium infection. The present sequence is a C. parvum coding
CC   sequence used in the exemplification of the invention
XX
SQ   Sequence 7334 BP; 2627 A; 1536 C; 1263 G; 1908 T; 0 U; 0 Other;

  Query Match             2.5%;  Score 200.4;  DB 6;  Length 7334;
  Best Local Similarity   65.5%;  Pred. No. 6.1e-32;
  Matches  308;  Conservative    0;  Mismatches  161;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461

Db       2686 ACTACGACAACAACAACTACTACTACGACAACAACAACAACAACTACTACTACTACTACA 2745

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521

Db       2746 ACCACAACAACTACAACCACGACAACTACAACCACAACCACAACTACCAAGAAACCAACA 2805

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581

Db       2806 ACAACAACAACAACAACAACTACTACTACTACAACAACAACAACAACAACTACTACTACT 2865

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641

Db       2866 ACTACTACTACTACTACTACTACTACTACTACTACCACAACAACCACAACCACAACCACA 2925

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701

Db       2926 ACTACCAAGAAACCAACAACAACAACAACAACAACAACAACAACTACTA-CTACAACCAC 2984

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761

Db       2985 GACAACAACAACCACGACAACCACAACCACAACTACAACTACCAAGAAACCAACAACTAC 3044

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821

Db       3045 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAACAACTACTAC 3104

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACA 7871

Db       3105 TACTACTACAACCACAACCACAACCACAACCGCAACCACAACTACCAAGA 3154


RESULT 15
AAF75507
ID   AAF75507 standard; DNA; 486 BP.
XX
AC   AAF75507;
XX
DT   09-MAY-2001  (first entry)
XX
DE   Polyglutamine tract coding sequence #2.
XX
KW   Animal model; polyglutamine tract; neurodegenerative disorder; HDJ1;
KW   heat shock protein 40; tetratricopeptide repeat protein 2; TPR2; stroke;
KW   myeloid leukaemia factor 1; MLF; human; fruit fly; Alzheimer's disease;
KW   Parkinson's disease; CJD; BSE; Huntington's disease; head trauma; cancer;
KW   ds.
XX
OS   Drosophila sp.
XX
PN   WO200112238-A1.
XX
PD   22-FEB-2001.
XX
PF   14-AUG-2000; 2000WO-US022496.
XX
PR   12-AUG-1999;   99US-0148933P.
PR   12-AUG-1999;   99US-0148934P.
PR   18-JAN-2000; 2000US-0177047P.
PR   19-MAY-2000; 2000US-0205720P.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Benzer S,  Kazemi-Esfarjani P;
XX
DR   WPI; 2001-147537/15.
XX
PT   Identifying genes or other compounds that modulate polyglutamine
PT   toxicity, useful for treating Alzheimer's disease, Parkinson's disease
PT   and Creutzfeldt-Jakob disease.
XX
PS   Disclosure; Fig 1A; 275pp; English.
XX
CC   The present invention describes a method of screening for genes which
CC   modulate polyglutamine toxicity using animal models with polyglutamine
CC   sequences that cause toxicity in the animal. The model is preferably
CC   Drosophila, and toxic polyglutamine sequences include the human and
CC   Drosophila heat shock protein 40/HDJ1, tetratricopeptide repeat protein 2
CC   (TPR2) and myeloid leukaemia factor 1 (MLF) genes. The model is useful
CC   for identifying treatments for neurodegenerative and proliferative
CC   disorders, including Alzheimer's disease, Parkinson's disease,
CC   Creutzfeldt-Jacob's disease (CJD), bovine spongiform encephalopathy
CC   (BSE), Huntington's disease, Machado-Joseph disease, spinocerebellar
CC   ataxias, dentatorubropallidoluysian atrophy, Kennedy's disease, stroke,
CC   head trauma and cancer
XX
SQ   Sequence 486 BP; 193 A; 164 C; 108 G; 21 T; 0 U; 0 Other;

  Query Match             2.4%;  Score 196.6;  DB 4;  Length 486;
  Best Local Similarity   73.9%;  Pred. No. 2e-31;
  Matches  264;  Conservative    0;  Mismatches   89;  Indels    4;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461

Db         75 ACAGCAGCAGCAACAGCAACAGCAGCAGCAACAACAGCAGCAGCAACAGCAACAGCAGCA 134

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521

Db        135 GCAACAACAGCAGCAGCAACAGCAACAGCAGCAGCAACAGCAGCAGCAACAGCAACAGCA 194

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581

Db        195 GCAGCAACAACAGCAGCAGCAACAGCAACAGCAGCAGCAACAACAGCAACAACAACAGCA 254

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641

Db        255 ACAGCAGCAGCAACAGCAGCAGCAACAGCAACAGCAGCAGCAACAACAGCAGCAGCAACA 314

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGA----CAACA 7697

Db        315 GCAACAGCAGCAGCAACAACAGCAGCAGCAACAGCAACAGCAGCAGCAACAGCAGCAGCA 374

Qy       7698 ACAAGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCA 7754

Db        375 ACAGCAACAGCAGCAGCAACAACAGCAGCTGCAACAGCAACAGCAGCAGCAACAACA 431


Search completed: March 30, 2004, 17:07:30
Job time : 1845 secs
```

```
OM nucleic - nucleic search, using sw model

Run on:         March 30, 2004, 16:30:14 ; Search time 351 Seconds
                                           (without alignments)
                                           12719.601 Million cell updates/sec

Title:          US-09-913-878A-1
Perfect score:  8045
Sequence:       1 ggtaccgggcccccctcga..........cgtattacaatcccagtcac 8045

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:      1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	220.8	2.7	5163	3	US-08-700-651-1	Sequence 1, Appli
	2	220.8	2.7	5163	3	US-08-928-361B-4	Sequence 4, Appli
	3	220.8	2.7	5163	4	US-09-588-995A-4	Sequence 4, Appli
	4	220.8	2.7	5318	3	US-08-700-651-2	Sequence 2, Appli
	5	220.8	2.7	5318	3	US-08-928-361B-3	Sequence 3, Appli
	6	220.8	2.7	5318	4	US-09-588-995A-3	Sequence 3, Appli
	7	200.4	2.5	5511	3	US-08-928-361B-2	Sequence 2, Appli
	8	200.4	2.5	5511	4	US-09-588-995A-2	Sequence 2, Appli
	9	200.4	2.5	7334	3	US-08-928-361B-1	Sequence 1, Appli
	10	200.4	2.5	7334	4	US-09-588-995A-1	Sequence 1, Appli
	11	194.6	2.4	867	4	US-09-216-393B-340	Sequence 340, App
	12	194.6	2.4	867	4	US-09-216-393B-342	Sequence 342, App
	13	194.6	2.4	1397	4	US-09-216-393B-343	Sequence 343, App
c	14	194.6	2.4	1397	4	US-09-216-393B-345	Sequence 345, App
	15	176.6	2.2	198	5	PCT-US95-10668-3	Sequence 3, Appli
	16	176.6	2.2	198	5	PCT-US95-10668-4	Sequence 4, Appli
	17	175	2.2	198	5	PCT-US95-10668-1	Sequence 1, Appli
	18	175	2.2	198	5	PCT-US95-10668-2	Sequence 2, Appli
	19	157.4	2.0	2032	4	US-09-241-581B-5	Sequence 5, Appli
	20	157.4	2.0	2032	4	US-08-265-428-5	Sequence 5, Appli
	21	157.4	2.0	2032	5	PCT-US95-07721-5	Sequence 5, Appli
	22	148	1.8	6558	4	US-09-491-356C-7	Sequence 7, Appli
	23	143.4	1.8	688	3	US-08-998-416-915	Sequence 915, App
	24	137.4	1.7	1086	1	US-08-415-751-27	Sequence 27, Appl
c	25	137.4	1.7	1086	1	US-08-415-751-28	Sequence 28, Appl
c	26	137.2	1.7	1086	1	US-08-415-751-47	Sequence 47, Appl
	27	135.8	1.7	1430	1	US-08-276-452A-25	Sequence 25, Appl
	28	135.8	1.7	1430	2	US-08-798-744-25	Sequence 25, Appl
	29	135	1.7	1690	1	US-08-276-452A-24	Sequence 24, Appl
	30	135	1.7	1690	2	US-08-798-744-24	Sequence 24, Appl
	31	132	1.6	1107	2	US-08-991-300-1	Sequence 1, Appli
	32	130	1.6	390	3	US-09-197-649-7	Sequence 7, Appli
	33	125.8	1.6	397	3	US-09-253-691-3	Sequence 3, Appli
	34	122.6	1.5	6794	4	US-09-491-356C-2	Sequence 2, Appli
c	35	120.4	1.5	1137	4	US-09-569-098A-104	Sequence 104, App
	36	119.8	1.5	2190	4	US-09-625-188-19	Sequence 19, Appl
c	37	119.2	1.5	1037	4	US-09-181-585-3	Sequence 3, Appli
c	38	119.2	1.5	1159	4	US-09-181-585-1	Sequence 1, Appli
c	39	119.2	1.5	1471	4	US-09-181-585-2	Sequence 2, Appli
	40	114.8	1.4	543	6	5273901-6	Patent No. 5273901
	41	114.8	1.4	2058	2	US-08-749-391-1	Sequence 1, Appli
	42	114.8	1.4	2058	3	US-09-390-200-1	Sequence 1, Appli
c	43	114.2	1.4	396	4	US-09-640-173-53	Sequence 53, Appl
c	44	114.2	1.4	396	4	US-09-713-550-53	Sequence 53, Appl
c	45	113.2	1.4	700	4	US-09-236-097-11	Sequence 11, Appl

ALIGNMENTS

```
RESULT 1
US-08-700-651-1
; Sequence 1, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
;  APPLICANT: PETERSEN, CAROLYN
;  APPLICANT: LEECH, JAMES
;  APPLICANT: NELSON, RICHARD, C.
;  APPLICANT: GUT, JIRI
;  TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
;  TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
;  TITLE OF INVENTION: INFECTIONS
;  FILE REFERENCE: 480.19-4(HV)
;  CURRENT APPLICATION NUMBER: US/08/700,651B
;  CURRENT FILING DATE: 1997-08-14
;  EARLIER APPLICATION NUMBER: 08/415,751
;  EARLIER FILING DATE: 1995-04-03
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 5163
;   TYPE: DNA
;   ORGANISM: Cryptosporidium parvum
US-08-700-651-1

  Query Match              2.7%;  Score 220.8;  DB 3;  Length 5163;
  Best Local Similarity   67.0%;  Pred. No. 1.5e-44;
  Matches 343;  Conservative    0;  Mismatches  167;  Indels    2;  Gaps    2;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
               ||  ||   |  |||||||||||||||||||||||||||||||||||||||||||||||
Db        668 ACTACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 727

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
              ||||||| || || || || || || || || || || || || || || ||||| |||
Db        728 ACAACAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACAACCACA 787

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
              ||||| ||||| ||||| ||||| |||||||| ||||| ||||||||||||| ||||||
Db        788 ACAACTACAACCACAACTACAACCACAACAACTACAACCACAACAACAACAACCACAACA 847

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
              || ||||| ||||| ||||| |||||||| ||||| |||||||| ||||| ||||| |||
Db        848 ACTACAACCACAACTACAACCACAACAACTACAACCACAACAACCACAACCACAACCACA 907

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
              || ||      |  | ||||||||||||| ||||||||||||||||||||| |  |   |
Db        908 ACTACCAAGAAACCAACAACAACAACTACAACAACAACAACAACAACAACA-ACTACTAC 966

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
```

```
Db    967 TACAACCACCACAACAACAACAACAACAACAACTACAACTACCAAGAAACCAACAACTAC 1026
Qy   7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db   1027 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAA-CAACTACTA 1085
Qy   7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881
Db   1086 CTACTACTACAACAACAACAACAACAACAACAACAACAACAACTACCACGAAACCAACAA 1145
Qy   7882 CGAGTGTATGAGTGACGAGTACAACATGAACC 7913
Db   1146 CAACAACAACAACTACTACTACTACTACAACC 1177
```

RESULT 2
US-08-928-361B-4
; Sequence 4, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5163 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-928-361B-4

Query Match 2.7%; Score 220.8; DB 3; Length 5163;
Best Local Similarity 67.0%; Pred. No. 1.5e-44;
Matches 343; Conservative 0; Mismatches 167; Indels 2; Gaps 2;

```
Qy   7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db    667 ACTACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 726
Qy   7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db    727 ACAACAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACAACCACA 786
Qy   7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db    787 ACAACTACAACCACAACTACAACCACAACAACTACAACCACAACAACAACAACCACAACA 846
Qy   7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db    847 ACTACAACCACAACTACAACCACAACAACTACAACCACAACAACCACAACCACAACCACA 906
Qy   7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db    907 ACTACCAAGAAACCAACAACAACAACTACAACAACAACAACAACAACAACA-ACTACTAC 965
Qy   7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db    966 TACAACCACCACAACAACAACAACAACAACAACTACAACTACCAAGAAACCAACAACTAC 1025
Qy   7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db   1026 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAA-CAACTACTA 1084
Qy   7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881
Db   1085 CTACTACTACAACAACAACAACAACAACAACAACAACAACAACAACTACCACGAAACCAACAA 1144
Qy   7882 CGAGTGTATGAGTGACGAGTACAACATGAACC 7913
Db   1145 CAACAACAACAACTACTACTACTACTACAACC 1176
```

RESULT 3
US-09-588-995A-4
; Sequence 4, Application US/09588995A
; Patent No. 6514697
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: BARNES, DEBRA A.
; APPLICANT: NELSON, RICHARD C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND
; TITLE OF INVENTION: ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-5
; CURRENT APPLICATION NUMBER: US/09/588,995A
; CURRENT FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 08/827,171
; PRIOR FILING DATE: 1997-03-27
; PRIOR APPLICATION NUMBER: 08/928,361
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: 08/700,651
; PRIOR FILING DATE: 1996-08-14
; PRIOR APPLICATION NUMBER: 08/415,751
; PRIOR FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 5163
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-09-588-995A-4

Query Match 2.7%; Score 220.8; DB 4; Length 5163;
Best Local Similarity 67.0%; Pred. No. 1.5e-44;
Matches 343; Conservative 0; Mismatches 167; Indels 2; Gaps 2;

```
Qy   7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db    667 ACTACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 726
Qy   7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db    727 ACAACAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACAACCACA 786
```

```
Qy     7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db      787 ACAACTACAACCACAACTACAACCACAACAACTACAACCACAACAACAACAACCACAACA 846

Qy     7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db      847 ACTACAACCACAACTACAACCACAACAACTACAACCACAACAACCACAACCACAACCACA 906

Qy     7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db      907 ACTACCAAGAAACCAACAACAACAACTACAACAACAACAACAACAACA-ACTACTAC 965

Qy     7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db      966 TACAACCACCACAACAACAACAACAACAACAACTACAACTACCAAGAAACCAACAACTAC 1025

Qy     7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db     1026 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAA-CAACTACTA 1084

Qy     7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881
Db     1085 CTACTACTACAACAACAACAACAACAACAACAACAACAACAACTACCACGAAACCAACAA 1144

Qy     7882 CGAGTGTATGAGTGACGAGTACAACATGAACC 7913
Db     1145 CAACAACAACAACTACTACTACTACTACAACC 1176
```

RESULT 4
US-08-700-651-2
; Sequence 2, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5318
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-2

Query Match 2.7%; Score 220.8; DB 3; Length 5318;
Best Local Similarity 67.0%; Pred. No. 1.5e-44;
Matches 343; Conservative 0; Mismatches 167; Indels 2; Gaps 2;

```
Qy     7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db      668 ACTACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 727

Qy     7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db      728 ACAACAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACAACCACA 787

Qy     7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db      788 ACAACTACAACCACAACTACAACCACAACAACTACAACCACAACAACAACAACCACAACA 847

Qy     7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db      848 ACTACAACCACAACTACAACCACAACAACTACAACCACAACAACCACAACCACAACCACA 907

Qy     7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db      908 ACTACCAAGAAACCAACAACAACAACTACAACAACAACAACAACAACA-ACTACTAC 966

Qy     7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db      967 TACAACCACCACAACAACAACAACAACAACAACTACAACTACCAAGAAACCAACAACTAC 1026

Qy     7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db     1027 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAA-CAACTACTA 1085

Qy     7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881
Db     1086 CTACTACTACAACAACAACAACAACAACAACAACAACAACAACTACCACGAAACCAACAA 1145

Qy     7882 CGAGTGTATGAGTGACGAGTACAACATGAACC 7913
Db     1146 CAACAACAACAACTACTACTACTACTACAACC 1177
```

RESULT 5
US-08-928-361B-3
; Sequence 3, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5318 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-928-361B-3

Query Match 2.7%; Score 220.8; DB 3; Length 5318;
Best Local Similarity 67.0%; Pred. No. 1.5e-44;
Matches 343; Conservative 0; Mismatches 167; Indels 2; Gaps 2;

```
Qy     7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
```

```
Db      667 ACTACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 726

Qy     7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521

Db      727 ACAACAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACAACCACA 786

Qy     7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581

Db      787 ACAACTACAACCACAACTACAACCACAACAACTACAACCACAACAACAACAACCACAACA 846

Qy     7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641

Db      847 ACTACAACCACAACTACAACCACAACAACTACAACCACAACAACCACAACCACAACCACA 906

Qy     7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701

Db      907 ACTACCAAGAAACCAACAACAACAACTACAACAACAACAACAACAACAACA-ACTACTAC 965

Qy     7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761

Db      966 TACAACCACCACAACAACAACAACAACAACAACTACAACTACCAAGAAACCAACAACTAC 1025

Qy     7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821

Db     1026 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAA-CAACTACTA 1084

Qy     7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881

Db     1085 CTACTACTACAACAACAACAACAACAACAACAACAACAACAACTACCACGAAACCAACAA 1144

Qy     7882 CGAGTGTATGAGTGACGAGTACAACATGAACC 7913

Db     1145 CAACAACAACAACTACTACTACTACTACAACC 1176
```

RESULT 6
US-09-588-995A-3
; Sequence 3, Application US/09588995A
; Patent No. 6514697
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: BARNES, DEBRA A.
; APPLICANT: NELSON, RICHARD C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND
; TITLE OF INVENTION: ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-5
; CURRENT APPLICATION NUMBER: US/09/588,995A
; CURRENT FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 08/827,171
; PRIOR FILING DATE: 1997-03-27
; PRIOR APPLICATION NUMBER: 08/928,361
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: 08/700,651
; PRIOR FILING DATE: 1996-08-14
; PRIOR APPLICATION NUMBER: 08/415,751
; PRIOR FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 5318
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-09-588-995A-3

```
  Query Match             2.7%;  Score 220.8;  DB 4;  Length 5318;
  Best Local Similarity   67.0%;  Pred. No. 1.5e-44;
  Matches  343;  Conservative    0;  Mismatches  167;  Indels    2;  Gaps    2;

Qy     7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
```


```
Db      667 ACTACGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 726

Qy     7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521

Db      727 ACAACAACGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACAACCACA 786

Qy     7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581

Db      787 ACAACTACAACCACAACTACAACCACAACAACTACAACCACAACAACAACAACCACAACA 846

Qy     7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641

Db      847 ACTACAACCACAACTACAACCACAACAACTACAACCACAACAACCACAACCACAACCACA 906

Qy     7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701

Db      907 ACTACCAAGAAACCAACAACAACAACTACAACAACAACAACAACAACAACA-ACTACTAC 965

Qy     7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761

Db      966 TACAACCACCACAACAACAACAACAACAACAACTACAACTACCAAGAAACCAACAACTAC 1025

Qy     7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821

Db     1026 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAA-CAACTACTA 1084

Qy     7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881

Db     1085 CTACTACTACAACAACAACAACAACAACAACAACAACAACAACTACCACGAAACCAACAA 1144

Qy     7882 CGAGTGTATGAGTGACGAGTACAACATGAACC 7913

Db     1145 CAACAACAACAACTACTACTACTACTACAACC 1176
```


RESULT 7
US-08-928-361B-2
; Sequence 2, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 2:

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5511 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-928-361B-2

  Query Match              2.5%;  Score 200.4;  DB 3;  Length 5511;
  Best Local Similarity   65.5%;  Pred. No. 1.8e-39;
  Matches  308;  Conservative    0;  Mismatches  161;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db       1018 ACTACGACAACAACAACTACTACTACGACAACAACAACAACAACTACTACTACTACTACA 1077

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db       1078 ACCACAACAACTACAACCACGACAACTACAACCACAACCACAACTACCAAGAAACCAACA 1137

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db       1138 ACAACAACAACAACAACAACTACTACTACTACAACAACAACAACAACAACTACTACTACT 1197

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db       1198 ACTACTACTACTACTACTACTACTACTACTACTACCACAACAACCACAACCACAACCACA 1257

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db       1258 ACTACCAAGAAACCAACAACAACAACAACAACAACAACAACAACTACTA-CTACAACCAC 1316

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db       1317 GACAACAACAACCACGACAACCACAACCACAACTACAACTACCAAGAAACCAACAACTAC 1376

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db       1377 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAACAACTACTAC 1436

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACA 7871
Db       1437 TACTACTACAACCACAACCACAACCACAACCGCAACCACAACTACCAAGA 1486

RESULT 8
US-09-588-995A-2
; Sequence 2, Application US/09588995A
; Patent No. 6514697
; GENERAL INFORMATION:
;  APPLICANT: PETERSEN, CAROLYN
;  APPLICANT:  BARNES, DEBRA A.
;  APPLICANT:  NELSON, RICHARD C.
;  APPLICANT:  GUT, JIRI
;  TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND
;  TITLE OF INVENTION:  ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM
;  TITLE OF INVENTION:  INFECTIONS
;  FILE REFERENCE: 480.19-5
;  CURRENT APPLICATION NUMBER: US/09/588,995A
;  CURRENT FILING DATE:  2000-06-06
;  PRIOR APPLICATION NUMBER: 08/827,171
;  PRIOR FILING DATE: 1997-03-27
;  PRIOR APPLICATION NUMBER: 08/928,361
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 08/700,651
;  PRIOR FILING DATE: 1996-08-14
;  PRIOR APPLICATION NUMBER: 08/415,751
;  PRIOR FILING DATE: 1995-04-03
;  NUMBER OF SEQ ID NOS: 115
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 5511
;   TYPE: DNA
;   ORGANISM: Cryptosporidium parvum
US-09-588-995A-2

  Query Match              2.5%;  Score 200.4;  DB 4;  Length 5511;
  Best Local Similarity   65.5%;  Pred. No. 1.8e-39;
  Matches  308;  Conservative    0;  Mismatches  161;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db       1018 ACTACGACAACAACAACTACTACTACGACAACAACAACAACAACTACTACTACTACTACA 1077

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db       1078 ACCACAACAACTACAACCACGACAACTACAACCACAACCACAACTACCAAGAAACCAACA 1137

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db       1138 ACAACAACAACAACAACAACTACTACTACTACAACAACAACAACAACAACTACTACTACT 1197

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db       1198 ACTACTACTACTACTACTACTACTACTACTACTACCACAACAACCACAACCACAACCACA 1257

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db       1258 ACTACCAAGAAACCAACAACAACAACAACAACAACAACAACAACTACTA-CTACAACCAC 1316

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db       1317 GACAACAACAACCACGACAACCACAACCACAACTACAACTACCAAGAAACCAACAACTAC 1376

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db       1377 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAACAACTACTAC 1436

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACA 7871
Db       1437 TACTACTACAACCACAACCACAACCACAACCGCAACCACAACTACCAAGA 1486

RESULT 9
US-08-928-361B-1
; Sequence 1, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
```

```
;     REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  650-324-1677
;     TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH:  7334 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-928-361B-1

  Query Match              2.5%;  Score 200.4;  DB 3;  Length 7334;
  Best Local Similarity   65.5%;  Pred. No. 2.1e-39;
  Matches  308;  Conservative    0;  Mismatches  161;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db       2686 ACTACGACAACAACAACTACTACTACGACAACAACAACAACAACTACTACTACTACTACA 2745

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db       2746 ACCACAACAACTACAACCACGACAACTACAACCACAACCACAACTACCAAGAAACCAACA 2805

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db       2806 ACAACAACAACAACAACAACTACTACTACTACAACAACAACAACAACAACTACTACTACT 2865

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db       2866 ACTACTACTACTACTACTACTACTACTACTACTACCACAACAACCACAACCACAACCACA 2925

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db       2926 ACTACCAAGAAACCAACAACAACAACAACAACAACAACAACAACTACTA-CTACAACCAC 2984

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db       2985 GACAACAACAACCACGACAACCACAACCACAACTACAACTACCAAGAAACCAACAACTAC 3044

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db       3045 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAACAACTACTAC 3104

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACA 7871
Db       3105 TACTACTACAACCACAACCACAACCACAACCGCAACCACAACTACCAAGA 3154

RESULT 10
US-09-588-995A-1
; Sequence 1, Application US/09588995A
; Patent No. 6514697
; GENERAL INFORMATION:
;  APPLICANT: PETERSEN, CAROLYN
;  APPLICANT:  BARNES, DEBRA A.
;  APPLICANT:  NELSON, RICHARD C.
;  APPLICANT:  GUT, JIRI
;  TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND
;  TITLE OF INVENTION:  ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM
;  TITLE OF INVENTION:  INFECTIONS
;  FILE REFERENCE: 480.19-5
;  CURRENT APPLICATION NUMBER: US/09/588,995A
;  CURRENT FILING DATE:  2000-06-06
;  PRIOR APPLICATION NUMBER: 08/827,171
;  PRIOR FILING DATE: 1997-03-27
;  PRIOR APPLICATION NUMBER: 08/928,361
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 08/700,651
;  PRIOR FILING DATE: 1996-08-14
;  PRIOR APPLICATION NUMBER: 08/415,751
;  PRIOR FILING DATE: 1995-04-03
;  NUMBER OF SEQ ID NOS: 115
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 7334
;   TYPE: DNA
;   ORGANISM: Cryptosporidium parvum
US-09-588-995A-1

  Query Match              2.5%;  Score 200.4;  DB 4;  Length 7334;
  Best Local Similarity   65.5%;  Pred. No. 2.1e-39;
  Matches  308;  Conservative    0;  Mismatches  161;  Indels    1;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db       2686 ACTACGACAACAACAACTACTACTACGACAACAACAACAACAACTACTACTACTACTACA 2745

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db       2746 ACCACAACAACTACAACCACGACAACTACAACCACAACCACAACTACCAAGAAACCAACA 2805

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db       2806 ACAACAACAACAACAACAACTACTACTACTACAACAACAACAACAACAACTACTACTACT 2865

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db       2866 ACTACTACTACTACTACTACTACTACTACTACTACCACAACAACCACAACCACAACCACA 2925

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db       2926 ACTACCAAGAAACCAACAACAACAACAACAACAACAACAACAACTACTA-CTACAACCAC 2984

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db       2985 GACAACAACAACCACGACAACCACAACCACAACTACAACTACCAAGAAACCAACAACTAC 3044

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db       3045 TACTACTACTACCACAACAACAACAACTACTACTACTACCACAACAACAACAACTACTAC 3104

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACA 7871
Db       3105 TACTACTACAACCACAACCACAACCACAACCGCAACCACAACTACCAAGA 3154

RESULT 11
US-09-216-393B-340
; Sequence 340, Application US/09216393B
; Patent No. 6514694
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THI
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393B
;  CURRENT FILING DATE:  1998-12-18
;  PRIOR APPLICATION NUMBER: 08/994,825
;  PRIOR FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 366
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 340
;   LENGTH: 867
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(867)
;   OTHER INFORMATION:
US-09-216-393B-340

  Query Match              2.4%;  Score 194.6;  DB 4;  Length 867;
  Best Local Similarity   79.6%;  Pred. No. 2e-38;
  Matches  230;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;
```

```
Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db        529 ACAACTACGACACCAACAACAACTACAACAACCACAACAACTACACCAACAACAACGACA 588

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db        589 ACAACCACAACAACTACACCAACAACAACGACAACAACCACAACAACTACACCAACAACA 648

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db        649 ACGACAACAACCACAACGCCAACTACAACGACATCTACGACAACCACTACGACTACCACA 708

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db        709 ACTACTACTACACCAACTACAACAACGACAACCACGGAACCAACAACTACAACAACAACC 768

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7690
Db        769 ACGGAACCAACCACAACTACAAGCACAACGACGACTACGACAACTACAA 817

RESULT 12
US-09-216-393B-342
; Sequence 342, Application US/09216393B
; Patent No. 6514694
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THER
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393B
;  CURRENT FILING DATE:  1998-12-18
;  PRIOR APPLICATION NUMBER: 08/994,825
;  PRIOR FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 366
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 342
;   LENGTH: 867
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
US-09-216-393B-342

  Query Match              2.4%;  Score 194.6;  DB 4;  Length 867;
  Best Local Similarity   79.6%;  Pred. No. 2e-38;
  Matches  230;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db        529 ACAACTACGACACCAACAACAACTACAACAACCACAACAACTACACCAACAACAACGACA 588

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db        589 ACAACCACAACAACTACACCAACAACAACGACAACAACCACAACAACTACACCAACAACA 648

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db        649 ACGACAACAACCACAACGCCAACTACAACGACATCTACGACAACCACTACGACTACCACA 708

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db        709 ACTACTACTACACCAACTACAACAACGACAACCACGGAACCAACAACTACAACAACAACC 768

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7690
Db        769 ACGGAACCAACCACAACTACAAGCACAACGACGACTACGACAACTACAA 817

RESULT 13
US-09-216-393B-343
; Sequence 343, Application US/09216393B
; Patent No. 6514694
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THER
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393B
;  CURRENT FILING DATE:  1998-12-18
;  PRIOR APPLICATION NUMBER: 08/994,825
;  PRIOR FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 366
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 343
;   LENGTH: 1397
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (238)..(1104)
;   OTHER INFORMATION:
US-09-216-393B-343

  Query Match              2.4%;  Score 194.6;  DB 4;  Length 1397;
  Best Local Similarity   79.6%;  Pred. No. 2.5e-38;
  Matches  230;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db        766 ACAACTACGACACCAACAACAACTACAACAACCACAACAACTACACCAACAACAACGACA 825

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db        826 ACAACCACAACAACTACACCAACAACAACGACAACAACCACAACAACTACACCAACAACA 885

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db        886 ACGACAACAACCACAACGCCAACTACAACGACATCTACGACAACCACTACGACTACCACA 945

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db        946 ACTACTACTACACCAACTACAACAACGACAACCACGGAACCAACAACTACAACAACAACC 1005

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7690
Db       1006 ACGGAACCAACCACAACTACAAGCACAACGACGACTACGACAACTACAA 1054

RESULT 14
US-09-216-393B-345/c
; Sequence 345, Application US/09216393B
; Patent No. 6514694
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THE
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393B
;  CURRENT FILING DATE:  1998-12-18
;  PRIOR APPLICATION NUMBER: 08/994,825
;  PRIOR FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 366
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 345
;   LENGTH: 1397
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
US-09-216-393B-345

  Query Match              2.4%;  Score 194.6;  DB 4;  Length 1397;
  Best Local Similarity   79.6%;  Pred. No. 2.5e-38;
  Matches  230;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db        632 ACAACTACGACACCAACAACAACTACAACAACCACAACAACTACACCAACAACAACGACA 573

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db        572 ACAACCACAACAACTACACCAACAACAACGACAACAACCACAACAACTACACCAACAACA 513
```

```
Qy        7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581

Db         512 ACGACAACAACCACAACGCCAACTACAACGACATCTACGACAACCACTACGACTACCACA 453

Qy        7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641

Db         452 ACTACTACTACACCAACTACAACAACGACAACCACGGAACCAACAACTACAACAACAACC 393

Qy        7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7690

Db         392 ACGGAACCAACCACAACTACAAGCACAACGACGACTACGACAACTACAA 344


RESULT 15
PCT-US95-10668-3
; Sequence 3, Application PC/TUS9510668
;  GENERAL INFORMATION:
;    APPLICANT:  James Eberwine
;    TITLE OF INVENTION:  A Method of Sequencing Proteins by
;    TITLE OF INVENTION:  Epitope Ordering and Protein
;    TITLE OF INVENTION:  Restriction Mapping
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jane Massey Licata, Esq.
;      STREET:  210 Lake Drive East, Suite 201
;      CITY:  Cherry Hill
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10668
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/294,133
;      FILING DATE:  August 22, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane.Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  PENN-0137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 779-2400
;      TELEFAX:  (609) 779-8488
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  198
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    ANTI-SENSE:  No
PCT-US95-10668-3

  Query Match             2.2%;  Score 176.6;  DB 5;  Length 198;
  Best Local Similarity   97.8%;  Pred. No. 2.9e-34;
  Matches  179;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        7412 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7471

Db           1 CAACAACAACAACAACAACAACAACAACAACAAGAATTCCAACAACAACAACAACAA 60

Qy        7472 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7531

Db          61 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 120

Qy        7532 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7591

Db         121 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 180

Qy        7592 CAA 7594

Db         181 CAA 183

Search completed: March 31, 2004, 02:06:48
Job time : 354 secs
```

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 16:28:24 ; Search time 12100 Seconds
(without alignments)
19854.648 Million cell updates/sec

Title: US-09-913-878A-1
Perfect score: 8045
Sequence: 1 ggtaccgggcccccctcga..........cgtattacaatcccagtcac 8045

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	275	3.4	869	14	CK159167	CK159167 FGAS04056
c	2	263	3.3	477	14	CA743784	CA743784 wri1s.pk0
	3	255.2	3.2	603	12	BJ366946	BJ366946 BJ366946
	4	254.2	3.2	356	12	BJ337970	BJ337970 BJ337970
	5	254	3.2	494	12	BJ360881	BJ360881 BJ360881
	6	250	3.1	394	9	AU060224	AU060224 AU060224
	7	247.4	3.1	1933	11	AY325173	AY325173 Rattus no
	8	245.8	3.1	969	11	AY383694	AY383694 Rattus no
	9	242.6	3.0	359	12	BJ366856	BJ366856 BJ366856
	10	242.6	3.0	627	12	BJ426316	BJ426316 BJ426316
	11	241.6	3.0	450	12	BJ371297	BJ371297 BJ371297
c	12	238.8	3.0	889	14	CK159613	CK159613 FGAS04107
	13	235.8	2.9	436	14	CB791539	CB791539 AMGNNUC:N
	14	234.8	2.9	256	14	CA702278	CA702278 wdk1c.pk0
c	15	234.8	2.9	732	28	BZ284533	BZ284533 CH230-385
	16	230.2	2.9	503	12	BM656118	BM656118 170006873
c	17	228.6	2.8	711	28	BZ098560	BZ098560 CH230-235
c	18	227	2.8	906	29	CNS02GJN	AL196412 Tetraodon
	19	225.8	2.8	827	29	CNS02156	AL176451 Tetraodon
	20	225.4	2.8	494	12	BM641696	BM641696 170006873
	21	225.4	2.8	546	13	C22974	C22974 C22974 Dict
c	22	225.4	2.8	681	29	CNS02EOD	AL193990 Tetraodon
	23	224.4	2.8	608	12	BJ330059	BJ330059 BJ330059
	24	223.2	2.8	551	12	BJ366220	BJ366220 BJ366220
	25	223.2	2.8	571	9	AU284550	AU284550 AU284550
	26	222.8	2.8	612	10	BE195101	BE195101 HVSMEh008
	27	221.8	2.8	970	29	CNS03H6V	AL243904 Tetraodon
	28	219.8	2.7	904	29	CNS03HA0	AL244017 Tetraodon
	29	218.8	2.7	818	28	BZ229437	BZ229437 CH230-401
	30	218.6	2.7	563	12	BJ371247	BJ371247 BJ371247
	31	218.2	2.7	652	13	BW243685	BW243685 BW243685
	32	216.4	2.7	429	9	AU052930	AU052930 AU052930
	33	215	2.7	617	9	AU033655	AU033655 AU033655
c	34	214.2	2.7	280	12	BJ377836	BJ377836 BJ377836
	35	214.2	2.7	745	13	BQ986791	BQ986791 QGF10H09.
	36	214.2	2.7	854	13	BX078177	BX078177 BX078177
	37	213.8	2.7	576	14	CD909608	CD909608 G468.113B
	38	213	2.6	878	28	BH153470	BH153470 ENTRE19TF
	39	212.8	2.6	410	12	BJ370460	BJ370460 BJ370460
	40	212.8	2.6	652	12	BP000526	BP000526 BP000526
	41	212.2	2.6	795	29	CNS03HRI	AL244647 Tetraodon
	42	211.8	2.6	317	12	BJ366180	BJ366180 BJ366180
	43	211.8	2.6	419	9	AU033391	AU033391 AU033391
	44	211.8	2.6	544	9	AU037837	AU037837 AU037837
	45	211.2	2.6	303	12	BJ367635	BJ367635 BJ367635

ALIGNMENTS

RESULT 1
CK159167/c
LOCUS CK159167 869 bp mRNA linear EST 05-DEC-2003
DEFINITION FGAS040564 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA, mRNA sequence.
ACCESSION CK159167
VERSION CK159167.1 GI:38985053
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 869)
AUTHORS Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D., Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A., Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D., Penniket,C., Roach,J.L. and Sarhan,F.
TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops
JOURNAL Unpublished (2003)
COMMENT Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033

```
            Email: fgas_ests@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [128,636].
            Plate: TaLt537  row: N  column: 23.
FEATURES             Location/Qualifiers
     source          1..869
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Wheat line PI 178383"
                     /db_xref="taxon:4565"
                     /lab_host="DH5 alpha"
                     /clone_lib="Triticum aestivum FGAS: TaLt5"
                     /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
                     subtractive hybridization) cDNA library from genotype
                     PI178383 cold hardened at 2 C for 21 days and 49 days
                     (equal amount of cDNA pooled together before subtraction,
                     tester) and subtracted against genotype Norstar cold
                     hardened at 2 C for 1 day (24 H)(driver). Modified Smart
                     cDNA (Clontech)priming and non-directional cloning"
ORIGIN

  Query Match             3.4%;  Score 275;  DB 14;  Length 869;
  Best Local Similarity   71.3%;  Pred. No. 6.5e-24;
  Matches  376;  Conservative    0;  Mismatches  150;  Indels    1;  Gaps    1;

Qy       7387 CAACTCTACCTGATCACATCGGAAGCAACAACAACAACAACAACAACAACAACAACA 7446
Db        745 CAACGACCACCAACAACACCAACAACAACAACAACAACACCAACAACAACAACAACA 686

Qy       7447 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7506
Db        685 ACAACAACAACGACAACCACCACCACAACCACAACAACAACAACAACGACAACAACA 626

Qy       7507 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7566
Db        625 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 566

Qy       7567 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7626
Db        565 ACAACAACAACAACGACAACAACAACAACAACAACAACAACAACAACAACAACAACA 506

Qy       7627 ACAACAACAACAACAACAACAGCAACAACAACAACAACAACAACAACAACAACAACA 7686
Db        505 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 446

Qy       7687 ACAA-GACAACAACAAGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGA 7745
Db        445 ACAACGACGACGACGACAACGACGACGACGACGACGACGACAACGACAACAACGACA 386

Qy       7746 GCGACCTCATAGCTTATGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAA 7805
Db        385 ACAACAACAACGACAACAACAACAACAACAACAACAACGACCACAACGACAACGACAACA 326

Qy       7806 CAACAATGAAGTTCAAAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCA 7865
Db        325 ACAACAACAACAACAACAACAACGACAACAACCACAACAACAACAACAACAACAACAACA 266

Qy       7866 CCCACAGCATAAACCTCGAGTGTATGAGTGACGAGTACAACATGAAC 7912
Db        265 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 219

RESULT 2
CA743784/c
LOCUS       CA743784                 477 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  wri1s.pk005.b14 wri1s Triticum aestivum cDNA clone wri1s.pk005.b14
            5' end, mRNA sequence.
ACCESSION   CA743784
VERSION     CA743784.1  GI:25559607
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
            Miao,G., Caraher,N. and Hanafey,M.K.
  TITLE     DuPont Wheat cDNA Sequence
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Scott V. Tingey
            Crop Genetics
            E. I. DuPont de Nemours and Company
            1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
            Tel: 302-631-2602
            Fax: 302-631-2607
            Email: Scott.V.Tingey@USA.dupont.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1..477
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4565"
                     /clone="wri1s.pk005.b14"
                     /tissue_type="leaf"
                     /lab_host="DH10B"
                     /clone_lib="wri1s"
                     /note="Vector: PGEM-T Easy; Site_1: SmaI; Riband
                     (susceptible) wheat leaves infected with Septoria tritici
                     strain A, 24 hours after infection, subtracted w/
                     comparable uninfected leaves"
ORIGIN

  Query Match             3.3%;  Score 263;  DB 14;  Length 477;
  Best Local Similarity   85.4%;  Pred. No. 2.2e-22;
  Matches  304;  Conservative    0;  Mismatches   48;  Indels    4;  Gaps    1;

Qy       7403 CATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7462
Db        399 CAGCNNNAACAACAACAGCAGCAACAACAACAACAACAACAACAACAACAACAGCAG 340

Qy       7463 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7522
Db        339 CAACAACAGCAGCAACAACAGCAGCAACAACAGCAGCAACAACAGCAGCAACAACAACAA 280

Qy       7523 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7582
Db        279 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 220

Qy       7583 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7642
Db        219 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 160

Qy       7643 CAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAA----GACAACAA 7698
Db        159 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAATCTTCACAACAC 100

Qy       7699 CAAGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCA 7754
Db         99 CAACAACAGCAGCAACAACAACAACAGCAACAACAACAACAACATCAACAACATCA 44

RESULT 3
BJ366946
LOCUS       BJ366946                 603 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ366946 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc40f20 5', mRNA sequence.
ACCESSION   BJ366946
VERSION     BJ366946.1  GI:19276248
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
```

```
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc40f20"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

  Query Match             3.2%;  Score 255.2;  DB 12;  Length 603;
  Best Local Similarity   80.4%;  Pred. No. 1.8e-21;
  Matches  307;  Conservative    0;  Mismatches   74;  Indels    1;  Gaps    1;

Qy       7350 TTGTTCTTTTGCCTTTTCGTTGAGTTGGAAAGTATATCAACTCTACCTGATCACATCGGA 7409
              || ||   ||    |  | |||  |   |   |     | || |  |     |     |  ||| |
Db        217 TTATTAAATTCTTTGATAGTTTACATTCATGTTGTAACGAACAAATTCAACAACAACAAG 276

Qy       7410 AGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7469
              | |||  ||||||||| |||||||||| || ||| |||||||||||||||||||||||||
Db        277 AACAAGAACAACAACTACAACAACAACTACTACAGCAACAACAACAACAACAACAACAAC 336

Qy       7470 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7529
              ||||||||||||||||||||||||||||||||||| || ||||||||||||||||| |
Db        337 AACAACAACAACAACAACAACAACAACAACAACAACAGCAGCAACAACAACAACAACAGC 396

Qy       7530 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7589
              ||||||||||||| | |||||||||||||||| |||| |||||| |||| ||||||| |||
Db        397 AACAACAACAACAGCCACAACAACAACAACAANAACAGCAACAAGAACAGCAACAANAAC 456

Qy       7590 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAGC 7649
              ||||  | ||  ||||||||||||||||||||||||||||||||||||||||||||| |
Db        457 AACANGAGCACNAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 516

Qy       7650 AACAACAACAACAACAACAACAACAACAACAACAACAACAA-GACAACAACAAGACAACT 7708
              ||||||||||||||||||||||||||||||||||||||  ||||||||||| |  |
Db        517 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 576

Qy       7709 ATCAGGTTTAACCTCAGTACCA 7730
              | ||    |||  ||  | ||
Db        577 AACAANAGNAACAACAACAACA 598


RESULT 4
BJ337970
LOCUS       BJ337970                 356 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ337970 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda59i17 5', mRNA sequence.
ACCESSION   BJ337970
VERSION     BJ337970.1  GI:19168100
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 356)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .356
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda59i17"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             3.2%;  Score 254.2;  DB 12;  Length 356;
  Best Local Similarity   90.0%;  Pred. No. 2.8e-21;
  Matches  271;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
              ||| |   | ||||||| || ||| ||||||||||||||||||||||||||||||||
Db         35 ACAACTACAACAACAACTACTACAGCAACAACAACAACAACAACAACAACAACAACAACA 94

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
              |||||||||||||||||||||||||| || ||||||||||||||||| ||||||||
Db         95 ACAACAACAACAACAACAACAACAACAGCAGCAACAACAACAACAACAGCAACAACA 154

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
              |||||| | |||||||||||||||| |||| |||||| |||| |||||| |||||||| |
Db        155 ACAACAGCCACAACAACAACAACAAGAACAGCAACAAGAACAGCAACAAGAACAACAAGA 214

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
               ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 GCACGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 274

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
              |||||| ||||||||||||||||||||||||||||||||||||||  |   |   |  |
Db        275 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAGCANNA 334

Qy       7702 G 7702
              |
Db        335 G 335


RESULT 5
BJ360881
LOCUS       BJ360881                 494 bp    mRNA    linear   EST 07-MAR-2002
DEFINITION  BJ360881 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc8k12 5', mRNA sequence.
ACCESSION   BJ360881
VERSION     BJ360881.1  GI:19260476
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .494
```

```
                    /organism="Dictyostelium discoideum"
                    /mol_type="mRNA"
                    /strain="AX4"
                    /db_xref="taxon:44689"
                    /clone="ddc8k12"
                    /sex="mat A"
                    /dev_stage="Culmination stage"
                    /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

  Query Match             3.2%;  Score 254;  DB 12;  Length 494;
  Best Local Similarity   79.5%;  Pred. No. 2.7e-21;
  Matches  310;  Conservative    0;  Mismatches   79;  Indels    1;  Gaps    1;

Qy       7341 TTTGATCTTTTGTTCTTTTGCCTTTTCGTTGAGTTGGAAAGTATATCAACTCTACCTGAT 7400

Db        103 TTAAATTCTTTGATAGTTTACATTCATGTTGTAACGAACAAATTCAACAACAACAAGAAC 162

Qy       7401 CACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7460

Db        163 AAGAACAACAACTACAACAACAACTACTACAGCAACAACAACAACAACAACAACAACAAC 222

Qy       7461 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7520

Db        223 AACAACAACAACAACAACAACAACAACAACAACAACAGCAGCAACAACAACAACAACAGC 282

Qy       7521 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7580

Db        283 AACAACAACAACAGCCACAACAACAACAACAAGAACAGCAACAAGAACAGCAACAAGAAC 342

Qy       7581 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7640

Db        343 AACANGAGCACGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 402

Qy       7641 AACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAA-GACAACAAC 7699

Db        403 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 462

Qy       7700 AAGACAACTATCAGGTTTAACCTCAGTACC 7729

Db        463 AACANNANCAACAACAACAACAACAACAAC 492


RESULT 6
AU060224
LOCUS       AU060224                 394 bp    mRNA    linear   EST 20-MAY-1999
DEFINITION  AU060224 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLA610, mRNA sequence.
ACCESSION   AU060224
VERSION     AU060224.1  GI:4881328
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1..394
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLA610"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

  Query Match             3.1%;  Score 250;  DB 9;  Length 394;
  Best Local Similarity   91.1%;  Pred. No. 8.8e-21;
  Matches  265;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461

Db        101 ACACCACCANCAACAACACCAACAACACCAACAACACCAACAACAACAACAACAACAACA 160

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521

Db        161 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 220

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581

Db        221 ACAACAACAATATCATCAACATCAGCAACAATATCATCAATATCAACAACAATACCATCA 280

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641

Db        281 ACCATCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 340

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGA 7692

Db        341 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 391


RESULT 7
AY325173
LOCUS       AY325173                1933 bp    mRNA    linear   HTC 26-JUL-2003
DEFINITION  Rattus norvegicus Aa2-050 mRNA, complete cds.
ACCESSION   AY325173
VERSION     AY325173.1  GI:33086523
KEYWORDS    HTC.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1933)
  AUTHORS   Xu,C.S., Li,W.Q., Li,Y.C., Yang,K.J., Yan,H.M., Chang,C.F.,
            Zhao,L.F., Ma,H., Wang,L., Wang,S.F., Han,H.P., Wang,G.P.,
            Chai,L.Q., Yuan,J.Y., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.
  TITLE     Liver regeneration after PH
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1933)
  AUTHORS   Xu,C.S., Li,W.Q., Li,Y.C., Yang,K.J., Yan,H.M., Chang,C.F.,
            Zhao,L.F., Ma,H., Wang,L., Wang,S.F., Han,H.P., Wang,G.P.,
            Chai,L.Q., Yuan,J.Y., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-2003) Henan Bioengineering Key Lab, Henan Normal
            University, NO. 148 Jianshe Road, Xinxiang City, Henan 453002, P.R.
            China
FEATURES             Location/Qualifiers
     source          1..1933
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
     CDS             41..1927
                     /note="liver regeneration-related protein LRRG146"
                     /codon_start=1
                     /product="Aa2-050"
                     /protein_id="AAP92574.1"
                     /db_xref="GI:33086524"
                     /translation="MAAGLDQGGPARCAGAGPGPPLLLRDSILAFVLQMTGSNEFKLN
                     QPPEDGISSVKFSPNTSQFLLVSSWDTSVRLYDVPANSMRLKYQHTGAVLDCAFYVEN
                     LVGTHDAPIRCVEYCPEVNVMVTGSWDQTVKLWDPRTPCNAGTFSQPEKGYVLSSIEG
                     RVAVEYLDPSPEVQKKKYAFKCHRLKENNIEQIYPVNAISLSQYPQHICHSNDGTTLA
                     IASSYMYEMDDTEHPEDGIFIRQVTDAETKPKSSRKSSVNICGLSQDGHRLAASSWRL
```

```
                  AVPGLWSSRIPVLTIPPSADTAKRQAPLDIATNCYSKFPDVHTAVTLVASMHKSRTGV
                  ITLLVFVTTYLMRSTIKEESLLLLVVKGGKVHHGGEGKSGSPKSQKHFLLALQFFIWH
                  NSNSNNNNNSNNNNSNNNNSNNNNNSNNSSSNNSNNSNNSNNSSSNSTSNNSNSNNS
                  NSNNSNNNNNNSNNSNSNNSNSNSNNSNNKNSNNNSNNNNSNSNSNNSNNSSNNSSS
                  NNSNSNNNNSNNSNNSNNSSNNSSSSNNSNNKNNSNNNNSNNNNNSNSSSSNNNSNSNN
                  NSNSSSSNNNSSSNNNSNNNSNNNSNSNNHHHHPGNLAAAPLICIGIFLLLQALKGLN
                  DTML"
ORIGIN

  Query Match            3.1%;  Score 247.4;  DB 11;  Length 1933;
  Best Local Similarity  69.1%;  Pred. No. 1e-20;
  Matches  353;  Conservative    0;  Mismatches  156;  Indels    2;  Gaps    1;

Qy       7402 ACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7461
Db       1299 ACAGCAGCAGCAACAACAGCAACAGCAACAACAGCAACAGCAACAACAGCAGCAGCAACA 1358

Qy       7462 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7521
Db       1359 GCACCAGCAACAACAGCAACAGCAACAACAGCAACAGCAACAACAGCAACAACAACAACA 1418

Qy       7522 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7581
Db       1419 ACAACAGCAACAACAGCAACAGCAACAACAGCAACAGCAACAGCAACAACAGCAATAACA 1478

Qy       7582 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7641
Db       1479 AAAACAGCAACAACAACAGCAACAACAACAACAGCAACAGCAACAGCAACAACAGCAACA 1538

Qy       7642 ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAACAA 7701
Db       1539 ACAGCAGCAACAATAGCAGCAGCAGCAACAACAGCAACAGCAACAACA--ACAGCAACAA 1596

Qy       7702 GACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGCTTA 7761
Db       1597 CAGCAACAACAGCAACAACAGCAGCAACAATAGCAGCAGCAGCAACAACAGCAACAACAA 1656

Qy       7762 TGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTTCAA 7821
Db       1657 AAACAACAGCAACAACAACAACAGCAACAACAACAACAACAGCAACAGCAGCAGCAGCAA 1716

Qy       7822 AGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAACCT 7881
Db       1717 CAACAACAGCAACAGCAACAACAACAGCAACAGCAGCAGCAGCAACAACAATAGCAGCAG 1776

Qy       7882 CGAGTGTATGAGTGACGAGTACAACATGAAC 7912
Db       1777 CAACAACAACAGCAACAACAACAGCAACAAC 1807

RESULT 8
AY383694
LOCUS       AY383694              969 bp    mRNA    linear   HTC 15-OCT-2003
DEFINITION  Rattus norvegicus LRRGT00039 mRNA, complete cds.
ACCESSION   AY383694
VERSION     AY383694.1  GI:37654293
KEYWORDS    HTC.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 969)
  AUTHORS   Xu,C.S., Chang,C.F., Han,H.P., Wang,G.P., Chai,L.Q., Yuan,J.Y.,
            Yang,K.J., Zhao,L.F., Ma,H., Wang,L., Wang,S.F., Xing,X.K.,
            Shen,G.M., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.
  TITLE     Liver regeneration after PH
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 969)
  AUTHORS   Xu,C.S., Chang,C.F., Han,H.P., Wang,G.P., Chai,L.Q., Yuan,J.Y.,
            Yang,K.J., Zhao,L.F., Ma,H., Wang,L., Wang,S.F., Xing,X.K.,
            Shen,G.M., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-2003) Henan Bioengineering Key Lab, Henan Normal
            University, No. 148 Jianshe Road, Xinxiang City, Henan 453002, P.R.
            China
FEATURES             Location/Qualifiers
     source          1..969
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     CDS             1..969
                     /note="liver regeneration-related"
                     /codon_start=1
                     /product="LRRGT00039"
                     /protein_id="AAQ96252.1"
                     /db_xref="GI:37654294"
                     /translation="MVVVMVVVVVVVVVVVVVVVVVVVVLGIESRDCSSNNSNNNNN
                     NSNNNNSNINNSNNSNNNNSNSNNSSNNNNNSNSNNSNNNSNINNSNNNNINNNNNNS
                     NNSNSNNNSNNNNSNSNNSNNSNNNNSNSNNSSNNNNNSNSNNSNNNSNNNNSNNNNI
                     NNNNNNSNSNNNNSNNSNNNNSNSNSNNSNSNKTAIATTATAITTTATTATAITTAIT
                     TTATATTATAITTTATTATAITTAITTTATATTATTAITTTATATTAAITTTVTVTTA
                     ITTATSTTATTTTTSITTTATAITTTATTATTTTATATTAIATKQQ"
ORIGIN

  Query Match            3.1%;  Score 245.8;  DB 11;  Length 969;
  Best Local Similarity  67.4%;  Pred. No. 2e-20;
  Matches  346;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy       7399 ATCACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7458
Db        287 ATAACAACAACAACAACAGCAACAACAGCAACAGCAATAACAACAGCAATAACAACAACA 346

Qy       7459 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7518
Db        347 GCAACAGCAACAACAGCAACAACAGCAATAACAACAACAGCAACAGCAACAACAGCAGCA 406

Qy       7519 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7578
Db        407 ATAACAACAACAACAGTAACAGTAACAACAGCAATAACAACAGCAATAACAACAACAGCA 466

Qy       7579 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7638
Db        467 ACAACAACAACATCAATAACAACAACAACAACAGCAACAGCAATAACAACAACAGCAACA 526

Qy       7639 ACAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAA 7698
Db        527 ACAGCAACAACAACAACAGCAACAGCAACAGCAACAACAGCAATAGCAACAAAACAGCAA 586

Qy       7699 CAAGACAACTATCAGGTTTAACCTCAGTACCATTTGTAACAGACCGAGCGACCTCATAGC 7758
Db        587 TAGCAACAACAGCAACAGCAATAACAACAACAGCAACAACAGCAACAGCAATAACAACAG 646

Qy       7759 TTATGGGAAAATAATACACGTCCAGAGATACATAACAAATACAGCAACAACAATGAAGTT 7818
Db        647 CAATAACAACAACAGCAACAGCAACAACAGCAACAGCAATAACAACAACAGCAACAACAG 706

Qy       7819 CAAAGAGTTTCTCGCCCGCCGTTCCACCACCACCAGCACCACCACCACCCACAGCATAAA 7878
Db        707 CAACAGCAATAACAACAGCAATAACAACAACAGCAACAGCAACAACAGCAACAACAGCAA 766

Qy       7879 CCTCGAGTGTATGAGTGACGAGTACAACATGAA 7911
Db        767 TAACAACAACAGCAACAGCAACAACAGCAGCAA 799

RESULT 9
BJ366856
LOCUS       BJ366856              359 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ366856 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc40b13 5', mRNA sequence.
ACCESSION   BJ366856
VERSION     BJ366856.1  GI:19276158
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
```

```
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .359
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc40b13"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

  Query Match             3.0%;  Score 242.6;  DB 12;  Length 359;
  Best Local Similarity   89.8%;  Pred. No. 7.1e-20;
  Matches  254;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       7410 AGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7469

Db          1 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 60

Qy       7470 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7529

Db         61 AACAACAACAACAACAACAACAACTACAACAACAACTACAACAACTACAACAACTACAAC 120

Qy       7530 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7589

Db        121 AACTACTACTACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 180

Qy       7590 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAGC 7649

Db        181 AACAACAACAACAACAACAACAACAACAACAACAACANATANAACAACANANAGAACAACNAC 240

Qy       7650 AACAACAACAACAACAACAACAACAACAACAACAACAACAAGA 7692

Db        241 NACTACNACNACTACAACAACCACNACTACAACNACTACAACA 283


RESULT 10
BJ426316
LOCUS       BJ426316                627 bp    mRNA    linear   EST 11-MAR-2002
DEFINITION  BJ426316 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv58g13 5', mRNA sequence.
ACCESSION   BJ426316
VERSION     BJ426316.1  GI:19343023
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv58g13"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

  Query Match             3.0%;  Score 242.6;  DB 12;  Length 627;
  Best Local Similarity   86.7%;  Pred. No. 5.8e-20;
  Matches  260;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy       7400 TCACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7459

Db        297 TCATGTTGTAACGAACAAATTCAACAACAACAAGAACAAGAACAACAACTACAACAACAA 356

Qy       7460 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7519

Db        357 CTACTACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 416

Qy       7520 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7579

Db        417 CAACAACAACAACAGCAGCAACAACAACAACAACAGCAACAACAACAACAGCCACAACAA 476

Qy       7580 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7639

Db        477 CAACAACAAGAACAGCAACAANAACAGCAACAANAACAACANGAGCACNAACAACAACGA 536

Qy       7640 CAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAAC 7699

Db        537 CAACAACANCAACAACNACAACAACAACAACAACAACNACANCAACNACACAACNACANC 596


RESULT 11
BJ371297
LOCUS       BJ371297                450 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ371297 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc57m07 5', mRNA sequence.
ACCESSION   BJ371297
VERSION     BJ371297.1  GI:19280680
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc57m07"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN
```

```
  Query Match             3.0%;  Score 241.6;  DB 12;  Length 450;
  Best Local Similarity   87.2%;  Pred. No. 8.6e-20;
  Matches  265;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       7399 ATCACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7458
              | |||| |   | |||||||| |||||||| |||||||| ||||||||||| ||||||||
Db        106 AGCACAACAACAACAACAACATCAACAACATCAACAACATCAACAACAACATCAACAACA 165

Qy       7459 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7518
              ||||||||| ||||||||||| ||||||||||| ||   |||||||||||| |||| |||
Db        166 ACAACAACATCAACAACAACATCAACAACAACAGCATATACAACAACAACAGCAACCACA 225

Qy       7519 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7578
              |||||| |||| |||||||||||||  ||||||||||||||||||||| |||||||||||
Db        226 ACAACAGCAACCACAACAACAACAATCACAACAACAACAACAACAACATCAACAACAACA 285

Qy       7579 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7638
              ||||| |||||||||||||| || |||||| ||||||||||||| || |||||| |||||
Db        286 ACAACCACAACAACAACAACCACCACAACAGCAACAACAACAACCACCACAACAGCAACA 345

Qy       7639 ACAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAACAA 7698
              ||||| ||||||||||||||||||||||||||||||||||||||||| |||| |  |  |
Db        346 ACAACCACAGCAACAACAACAACAACAACAACAACAACAACAACAACCACAACAACAACA 405

Qy       7699 CAAG 7702
                ||
Db        406 ACAG 409
```

RESULT 12
CK159613/c
LOCUS   CK159613   889 bp   mRNA   linear   EST 05-DEC-2003
DEFINITION FGAS041079 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA, mRNA sequence.
ACCESSION   CK159613
VERSION   CK159613.1 GI:38985955
KEYWORDS   EST.
SOURCE   Triticum aestivum (bread wheat)
  ORGANISM Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 889)
  AUTHORS   Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,
            Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
            Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D.,
            Penniket,C., Roach,J.L. and Sarhan,F.
  TITLE   Functional Genomics of Abiotic Stress In Wheat and Canola Crops
  JOURNAL   Unpublished (2003)
COMMENT   Contact: Wm L Crosby
            Bioinformatics
            University of Saskatchewan, Department of Computer Science
            1C101 Engineering Building, 57 Campus Drive, Saskatoon,
            Saskatchewan, S7N 5A9, Canada
            Tel: 306 966 1769
            Fax: 306 966 2033
            Email: fgas_ests@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [123,668].
            Plate: TaLt539 row: N column: 07.
FEATURES   Location/Qualifiers
     source   1. .889
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Wheat line PI 178383"
                     /db_xref="taxon:4565"
                     /lab_host="DH5 alpha"
                     /clone_lib="Triticum aestivum FGAS: TaLt5"
                     /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
                     subtractive hybridization) cDNA library from genotype
                     PI178383 cold hardened at 2 C for 21 days and 49 days
                     (equal amount of cDNA pooled together before subtraction,
                     tester) and subtracted against genotype Norstar cold
                     hardened at 2 C for 1 day (24 H)(driver). Modified Smart
                     cDNA (Clontech)priming and non-directional cloning"
ORIGIN

```
  Query Match             3.0%;  Score 238.8;  DB 14;  Length 889;
  Best Local Similarity   85.0%;  Pred. No. 1.5e-19;
  Matches  267;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy       7379 AAGTATATCAACTCTACCTGATCACATCGGAAGCAACAACAACAACAACAACAACAACAA 7438
              ||  | | ||||   |   ||  ||| |   | ||||||||||||||||||| || || |
Db        487 AACAACAACAACAACAAGAGACAACAACAACAACAACAACAACAACAACAACGACGACGA 428

Qy       7439 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7498
              | ||||| || || || || || || |||||||| |||||||| |||||||||||||| |
Db        427 CGACAACGACGACGACGACGACGACGACAACAACGACAACAACGACAACAACAACAACGA 368

Qy       7499 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7558
              ||||||||||||||||||||| ||| || ||||| ||||| |||||||||||||||||||
Db        367 CAACAACAACAACAACAACAATAACGACCACAACGACAACGACAACAACAACAACAACAA 308

Qy       7559 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 7618
              |||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CAACAACAATGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 248

Qy       7619 CAACAACAACAACAACAACAACAACAACAGCAACAACAACAACAACAACAACAACAACAA 7678
              ||||||||||||||||||||||||||||| ||||||||||||| || ||||| || ||||
Db        247 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACGACGACAACGACGACAA 188

Qy       7679 CAACAACAACAAGA 7692
              | | |||||| | |
Db        187 CGATAACAACGACA 174
```

RESULT 13
CB791539
LOCUS   CB791539   436 bp   mRNA   linear   EST 16-MAY-2003
DEFINITION AMGNNUC:NRHY7-00028-F7-A nrhy7 (10850) Rattus norvegicus cDNA clone nrhy7-00028-f7 5', mRNA sequence.
ACCESSION   CB791539
VERSION   CB791539.1 GI:29879932
KEYWORDS   EST.
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 436)
  AUTHORS   Amgen EST Program.
  TITLE   Amgen Rat EST Program
  JOURNAL   Unpublished (2003)
COMMENT   Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00028 row: f column: 7.
FEATURES   Location/Qualifiers
     source   1. .436
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="nrhy7-00028-f7"
                     /clone_lib="nrhy7 (10850)"
                     /note="Vector: pSPORT1; Site_1: Sall; Site_2: Not1; W Rat
                     hypothalamus adult female Wistar rat avg. Insert size 2.3
                     kb fraction 6 and 7"
ORIGIN

```
  Query Match             2.9%;  Score 235.8;  DB 14;  Length 436;
  Best Local Similarity   87.5%;  Pred. No. 4.3e-19;
  Matches  258;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;
```

```
Qy       7398 GATCACATCGGAAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7457
Db        136 GACAACAACGACAATGACAACGACAACGACAACAACAACGACGACGACAACAACAACAAC 195

Qy       7458 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7517
Db        196 AACAACAACAACGACAACAACAACAACAACAACGAAAACAACAACAACAACGAAAACGAC 255

Qy       7518 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7577
Db        256 AACAACAACAACAACAATGACGACAACAACAACAACAGCAACAACGACAACAACAACAAC 315

Qy       7578 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7637
Db        316 AACAACGACAACAACAACAACAACGAAAACAACAACAACAACGAAAACGACAACAACAAC 375

Qy       7638 AACAACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAAGA 7692
Db        376 AACAACAATGACGACAACAACAACAACAGCAACAACGACAACAACAACAACAACA 430
```

```
RESULT 14
CA702278
LOCUS       CA702278                 256 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  wdk1c.pk005.a18 wdk1c Triticum aestivum cDNA clone wdk1c.pk005.a18
            5' end, mRNA sequence.
ACCESSION   CA702278
VERSION     CA702278.1  GI:25424071
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 256)
  AUTHORS   Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
            Miao,G., Caraher,N. and Hanafey,M.K.
  TITLE     DuPont Wheat cDNA Sequence
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Scott V. Tingey
            Crop Genetics
            E. I. DuPont de Nemours and Company
            1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
            Tel: 302-631-2602
            Fax: 302-631-2607
            Email: Scott.V.Tingey@USA.dupont.com
            Seq primer: M13.
FEATURES             Location/Qualifiers
     source          1..256
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4565"
                     /clone="wdk1c.pk005.a18"
                     /tissue_type="kernel"
                     /clone_lib="wdk1c"
                     /note="Vector: pBluescript SK+; Wheat (Triticum aestivum
                     L.) developing kernel, 3 days after anthesis."
ORIGIN

  Query Match             2.9%;  Score 234.8;  DB 14;  Length 256;
  Best Local Similarity   96.4%;  Pred. No. 6.9e-19;
  Matches  239;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       7413 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7472
Db          1 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 60

Qy       7473 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7532
Db         61 AACAACAACAACAACAACANCAACAACAACAACAACAACAACAACAACANCAACAACAAC 120

Qy       7533 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7592
Db        121 AACAACCACAACAACGACAACAACGACAACCACAACAACGACAACAACAACAACAACAAC 180

Qy       7593 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAGCAAC 7652
Db        181 AACAACAACCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 240

Qy       7653 AACAACAA 7660
Db        241 AACAACAA 248
```

```
RESULT 15
BZ284533/c
LOCUS       BZ284533                 732 bp    DNA     linear   GSS 15-OCT-2002
DEFINITION  CH230-385E7.TVB CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-385E7, genomic survey sequence.
ACCESSION   BZ284533
VERSION     BZ284533.1  GI:24014662
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 732)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-385E7.TJC
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 385  row: E  column: 7
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..732
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-385E7"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             2.9%;  Score 234.8;  DB 28;  Length 732;
  Best Local Similarity   88.8%;  Pred. No. 4.7e-19;
  Matches  254;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       7410 AGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7469
Db        374 AACAACAACAACAGCACCAACAACAACAACAACAACAGCAGCAACAACAACAACAACAGC 315

Qy       7470 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7529
```

```
Db        314 AGCAGCAACAACAACAACAACAGCAACAACAACAACAACAACAGCAGCAACAACAACAAC 255

Qy       7530 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 7589
              | || || ||||||||||||||||||||| || || || ||||||||||| || || ||||
Db        254 AGCAGCAGCAACAACAACAACAACAACAGCAGCAGCAGCAACAACAACAGCAGCAGCAAC 195

Qy       7590 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAGC 7649
              ||||||||||||| ||||||||||||||||||||||| || |||||||||||||| ||||
Db        194 AACAACAACAACAGCAACAACAACAACAACAACAACAGCAGCAACAACAACAACAGCAGC 135

Qy       7650 AACAACAACAACAACAACAACAACAACAACAACAACAACAAGACAA 7695
              | |||||||||||||||||||| || || || ||||||||| | ||
Db        134 AGCAACAACAACAACAACAACAGCAGCAGCAGCAACAACAACAAAA 89
```

Search completed: March 31, 2004, 02:00:38
Job time : 12105 secs

GenCore version 5.1.6


```
OM protein - protein search, using sw model

Run on:         March 29, 2004, 08:19:26 ; Search time 66 Seconds
                                           (without alignments)
                                           6001.994 Million cell updates/sec

Title:          US-09-913-878A-2
Perfect score:  7397
Sequence:       1 MNPITPRKRNSPVEEIINRL..........YEVLGDDDFDGIGFTGNGDY 1402

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_29Jan04:*
               1:  geneseqp1980s:*
               2:  geneseqp1990s:*
               3:  geneseqp2000s:*
               4:  geneseqp2001s:*
               5:  geneseqp2002s:*
               6:  geneseqp2003as:*
               7:  geneseqp2003bs:*
               8:  geneseqp2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7397	100.0	1402	3	AAB13956	Aab13956 Neurospor
2	328	4.4	1196	4	AAG67235	Aag67235 Amino aci
3	318	4.3	836	3	AAB28532	Aab28532 Rice RNA-
4	297	4.0	1114	4	AAU10006	Aau10006 Tomato RN
5	297	4.0	1114	4	AAE00897	Aae00897 Tomato C-
6	292	3.9	1164	3	AAB28531	Aab28531 Maize RNA
7	280	3.8	1125	3	AAB28533	Aab28533 Soybean R
8	267.5	3.6	1108	3	AAB28530	Aab28530 Maize RNA
9	236	3.2	417	3	AAB28529	Aab28529 Maize par
10	182	2.5	1219	3	AAY78795	Aay78795 Human ant
11	179.5	2.4	966	4	AAM38679	Aam38679 Human pol
12	176	2.4	1555	6	ABJ25640	Abj25640 Aspergill
13	176	2.4	1832	6	ABJ26240	Abj26240 Aspergill
14	173.5	2.3	2271	4	ABB65616	Abb65616 Drosophil
15	168	2.3	1226	6	ABR52986	Abr52986 Protein s
16	165.5	2.2	2291	5	AAU75351	Aau75351 Drosophil
17	164.5	2.2	4386	4	ABG07375	Abg07375 Novel hum
18	163.5	2.2	4274	4	ABG00972	Abg00972 Novel hum
19	163.5	2.2	4397	4	ABG21944	Abg21944 Novel hum
20	160.5	2.2	1024	5	ABB57200	Abb57200 Mouse isc
21	159.5	2.2	1596	7	AAE37929	Aae37929 Human CGD
22	158.5	2.1	1736	5	AAU84308	Aau84308 Human end
23	157	2.1	1026	6	ABO14710	Abo14710 Novel hum
24	156.5	2.1	1057	6	ABO14709	Abo14709 Novel hum
25	156	2.1	1013	4	AAM38678	Aam38678 Human pol
26	154	2.1	1025	4	AAM38680	Aam38680 Human pol
27	153.5	2.1	1828	4	ABB50164	Abb50164 Human tra
28	153	2.1	571	5	ABP65975	Abp65975 Bifidobac
29	153	2.1	2001	6	ABU70405	Abu70405 Human adi
30	153	2.1	5373	4	AAU14603	Aau14603 Novel bon
31	153	2.1	5447	4	AAU14697	Aau14697 Novel bon
32	152	2.1	1127	4	ABB58310	Abb58310 Drosophil
33	152	2.1	2194	4	AAM40114	Aam40114 Human pol
34	151.5	2.0	218	4	AAE00898	Aae00898 Tomato RN
35	151	2.0	4873	6	ABO14747	Abo14747 Novel hum
36	150	2.0	5909	4	ABG23295	Abg23295 Novel hum
37	150	2.0	6619	4	ABG23329	Abg23329 Novel hum
38	149	2.0	2462	6	ABU61813	Abu61813 Mouse nuc
39	148.5	2.0	1418	6	ABR53853	Abr53853 Protein s
40	148.5	2.0	1457	3	AAY93418	Aay93418 Amino aci
41	148.5	2.0	1605	4	ABB70375	Abb70375 Drosophil
42	146	2.0	530	6	ADA55193	Ada55193 Human pro
43	146	2.0	1493	4	ABB69228	Abb69228 Drosophil
44	145.5	2.0	1400	6	ABP96237	Abp96237 Human nuc
45	145	2.0	1920	4	ABB64441	Abb64441 Drosophil

ALIGNMENTS

```
RESULT 1
AAB13956
ID   AAB13956 standard; protein; 1402 AA.
XX
AC   AAB13956;
XX
DT   16-NOV-2000  (first entry)
XX
DE   Neurospora crassa QDE-1 protein.
XX
KW   Gene silencing; quelling deficient; qde-1; al-1.
XX
OS   Neurospora crassa.
XX
PN   WO200050581-A2.
XX
PD   31-AUG-2000.
XX
PF   16-FEB-2000; 2000WO-IT000048.
XX
PR   22-FEB-1999;   99IT-RM000117.
XX
PA   (UYRO-) UNIV ROMA LA SAPIENZA.
XX
PI   Macino G,  Cogoni C;
XX
DR   WPI; 2000-579171/54.
DR   N-PSDB; AAA65171.
XX
PT   Novel polynucleotide encoding a polypeptide which has a silencing
PT   activity and comprising a RNA-dependent RNA polymerase domain.
XX
PS   Claim 21; Fig 4; 48pp; English.
XX
CC   The present sequence is the Neurospora crassa QDE-1 protein. This protein
CC   has gene silencing activity. The qde-1 gene was isolated by mutational
CC   analysis of an al-1 transgenic strain. This strain had an albino
CC   phenotype resulting from post-transcriptional silencing of the endogenous
CC   al-1 gene. Reversion of this phenotype indicated a mutation in a
CC   silencing gene. The silencing gene, qde-1, could then be isolated.
CC   Modulation of qde-1 expression may be used to inactivate genes and to
CC   silence suppression of genes
XX
SQ   Sequence 1402 AA;

  Query Match             100.0%;  Score 7397;  DB 3;  Length 1402;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1402;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MNPITPRKRNSPVEEIINRLNNDYNLGLQCVADTTLTPHRRKELAESDEDFGRHDKIYRA 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNPITPRKRNSPVEEIINRLNNDYNLGLQCVADTTLTPHRRKELAESDEDFGRHDKIYRA 60

Qy         61 LNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLPWSKEPPRAATAGQQWALQ 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLPWSKEPPRAATAGQQWALQ 120

Qy        121 TVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLT 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLT 180

Qy        181 RSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQCTKRAKGKLSDNVAAAAAPPVPI 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQCTKRAKGKLSDNVAAAAAPPVPI 240

Qy        241 ASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSSFE 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSSFE 300

Qy        301 APPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQPLSSSSQGETSFSTYYESFPS 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 APPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQPLSSSSQGETSFSTYYESFPS 360

Qy        361 SGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEV 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEV 420

Qy        421 TRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFVTA 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFVTA 480

Qy        481 MTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILI 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 MTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILI 540

Qy        541 PSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLR 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 PSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLR 600

Qy        601 AEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNN 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNN 660

Qy        661 TWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAK 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAK 720

Qy        721 RIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTL 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTL 780

Qy        781 EVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVE 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 EVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVE 840

Qy        841 FRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQK 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 FRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQK 900

Qy        901 RKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLV 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLV 960

Qy        961 ARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPE 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPE 1020

Qy       1021 IVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPN 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 IVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPN 1080

Qy       1081 FLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGA 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 FLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGA 1140

Qy       1141 LSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWD 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWD 1200

Qy       1201 PDLASYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQ 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PDLASYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQ 1260

Qy       1261 VYEKWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHKSPKF 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 VYEKWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHKSPKF 1320

Qy       1321 VWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGDGSEYPDP 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 VWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGDGSEYPDP 1380

Qy       1381 EVYEVLGDDDFDGIGFTGNGDY 1402
                ||||||||||||||||||||||
Db       1381 EVYEVLGDDDFDGIGFTGNGDY 1402
```

```
RESULT 2
AAG67235
ID   AAG67235 standard; protein; 1196 AA.
XX
AC   AAG67235;
XX
DT   13-NOV-2001  (first entry)
XX
DE   Amino acid sequence of a plant SGS2 polypeptide.
XX
KW   SGS2; RNA-dependent RNA polymerase; transgene silencing;
KW   transgene stability; crop plant; viral resistance.
XX
OS   Arabidopsis thaliana.
XX
PN   FR2804128-A1.
XX
PD   27-JUL-2001.
XX
PF   26-JAN-2000; 2000FR-00001007.
XX
PR   26-JAN-2000; 2000FR-00001007.
XX
PA   (RHOB-) RHOBIO.
PA   (INRG ) INST NAT RECH AGRONOMIQUE.
XX
PI   Beclin C,  Elmayan T,  Mourrain P,  Vaucheret H;
XX
DR   WPI; 2001-543303/61.
DR   N-PSDB; AAH77704.
XX
PT   A new plant SGS2 gene involved in encoding an RNA-dependent RNA
PT   polymerase and in transgene silencing, increases transgene stability and
PT   expression in transgenic plants when it is inactivated.
XX
PS   Claim 24; Fig 1; 46pp; French.
XX
CC   The present sequence represents a plant SGS2 polypeptide. SGS2 is an RNA-
CC   dependent RNA polymerase and is involved in transgene silencing.
CC   Inactivation of SGS2 is used to increase transgene stability and
CC   expression in plants, particularly crop plants, especially maize, corn,
CC   barley, sorghum, soya, sugar cane, beet, tobacco or cotton plants.
CC   Overexpression of SGS2 can be used to increase resistance to viral
CC   infection in plants
XX
SQ   Sequence 1196 AA;

  Query Match             4.4%;  Score 328;  DB 4;  Length 1196;
```

```
 Best Local Similarity   22.7%;  Pred. No. 3.6e-17;
 Matches  197;  Conservative  126;  Mismatches  312;  Indels  234;  Gaps   42;

Qy        535 FFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPL 594
Db        466 FVAPIVKDLTSSSFSQKTYVFKRVKSI------LTDG-FKLCGRKY-SFLAFSAN----- 512

Qy        595 REFQLRAEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWL 654
Db        513 ---QLR--------DRSAWFFAEDG--------KTR----------------VSDIKTWM 537

Qy        655 LQL-DNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGR 713
Db        538 GKFKDKNV----AKCAARMGLCFSSTYATVDVMPHEV---DTEVPDIERNGYVFSDGIGT 590

Qy        714 MSRSVAKRIRDVLGLGDV---PSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQ----- 765
Db        591 ITPDLADEVMEKLKL-DVHYSPCAYQIRYAGFKGV-------------VARWPSKSDGIR 636

Qy        766 ---RKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLIN 822
Db        637 LALRDSMKKFFSKH-TILEICS-WTRFQPGFLNRQIITLLS---------VLGVPDEIFW 685

Qy        823 DLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSG 882
Db        686 DMQ---ESMLYKLNRILDDTDVAFEVLTASCA---------------EQGNTAAIMLSAG 727

Qy        883 FDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFS-- 940
Db        728 FKPKTEPHLRGMLSSVRIAQLWGLREKSRIFVTSGRWLMGCLDEAGILEHGQCFIQVSKP 787

Qy        941 ---------SKFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVII 990
Db        788 SIENCFSKHGSRFKETKTDLEVVKGY-VAIAKNPCLHPGDVRILEAVDVPQLHHMYDCLI 846

Qy        991 FSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRY--LKKDKTTFK 1048
Db        847 FPQKGDRPHTNEASGSDLDGDLYFVAWDQKLI----------PPNRKSYPAMHYDAAEEK 896

Qy       1049 QLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTN----YKERLCYINNSVSNKPA 1104
Db        897 SL------GRAVNHQDIIDFFARN----LANEQLGTICNAHVVHADRSEY---GAMDEEC 943

Qy       1105 IILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPD-PMYKSDSWLGR------- 1156
Db        944 LLLAELAATAVDFPKTGKIVSMP--FHLKPKLYPDFMGKEDYQTYKSNKILGRLYRRVKE 1001

Qy       1157 ------------GEPTHI-----IDYLKFSIARP-------AIDKELEAFHNAMKAAKD 1191
Db       1002 VYDEDAEASSEESTDPSAIPYDAVLEIPGFEDLIPEAWGHKCLYDGQLIGLLGQYKVQKE 1061

Qy       1192 TE-DGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDS 1250
Db       1062 EEIVTGHIWS---MPKYTSKKQGELKER--------LKHSYNSLKKEF-RKVFEETIPDH 1109

Qy       1251 KDPYPVRVNQVYEK----WCAIT--PEAMDKSGANYDSKVIRLLELSFLADREMNTWALL 1304
Db       1110 ENLSEEEKNILYEKKASAWYHVTYHPEWVKKS-----------LELQ---DPDESSHAAM 1155

Qy       1305 RASTAFKLYYHKSPKFVWQMAGRQLAYIK 1333
Db       1156 LS-------------FAW-IAADYLARIK 1170
```

RESULT 3
AAB28532
ID   AAB28532 standard; protein; 836 AA.
XX
AC   AAB28532;
XX
DT   07-FEB-2001 (first entry)
XX
DE   Rice RNA-directed RNA polymerase, SEQ ID NO: 10.
XX
KW   Rice; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW   plant virus resistance; plant breeding.
XX
OS   Oryza sativa.
XX
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   N-PSDB; AAC63741.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 10; Page 46-49; 62pp; English.
XX
CC   The present sequence is part of a plant RNA-directed RNA polymerase
CC   protein. Polynucleotides encoding RNA-directed RNA polymerases were
CC   isolated from plant cDNA libraries. They are useful as probes for
CC   genetically and physically mapping genes, and as markers for traits
CC   linked to those genes. They are useful for controlling gene expression
CC   and provide mechanisms to engineer plant virus resistance. They are also
CC   useful for plant breeding to develop lines with desired phenotypes
XX
SQ   Sequence 836 AA;

```
  Query Match             4.3%;  Score 318;  DB 3;  Length 836;
  Best Local Similarity   21.1%;  Pred. No. 1.3e-16;
  Matches  225;  Conservative  162;  Mismatches  352;  Indels  328;  Gaps   52;

Qy        376 LARSEESARSQVQVHAPVVAARLRNIW-PKFPKWLHEAPLAVAWEVTRLFMHC------- 427
Db          1 MXRSEEN----------VNVAALRDFWGDKFP----------------VFDACGRLKKAL 34

Qy        428 -KVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPP-----------N 474
Db         35 NRVARNPKLLCSKVGDDHAEVRSVVITPTXAYCL-----------PPEVERSNRVLRHYH 83

Qy        475 DVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDR 534
Db         84 EVADRFLRVTFMDEGMQVLNNNVL-----NSFTAPI----VKDLM------------SN 121

Qy        535 FFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPL 594
Db        122 FFQ-----------------QKTTVYKRVRMLLTEGFH-MCGRKY-SFLAFSSN----- 156

Qy        595 REFQLRAEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWL 654
Db        157 ---QLR--------DKSAWFFAE--------------------DRKT--TVEAIRKW- 180

Qy        655 LQLDNNTWQPHLKLF---SRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGV 711
Db        181 --MDGS----QVRMFEDAARMGQCFSSTYATVTMRPDEVDESFDDVVH---NEYIFSDGI 231

Qy        712 GRMSRSVAKRIRDVLGLGD-VPSAVQGRFGSAKGMWVIDV-DDTGDEDWIETYPSQRKWE 769
Db        232 GKITPDLALEVAERLQLTDNPPSAYQIRFAGFKG--VIAVWQGHGDGTRLFLRPSMRKFE 289

Qy        770 CDFVDKHQRTLEVRSVAS--ELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQ 827
```

```
Db        290 SNHL--------VLGVVSWTKFQPGFLNRQIIILLSS---------LNVPDSIFWQMQET 332

Qy        828 FSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKK 887

Db        333 MLSN---LNNILSDRDVAFEVLTTSCA---------------DDGNTAALMLSAGFEPRT 374

Qy        888 QKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEE------------NEV 935

Db        375 EPHLKAMLLAIRSAQLQDLLEKARIFVPKGRWLMGCLDELGVLEQGQCFIRATVPSLNSY 434

Qy        936 HVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKG 995

Db        435 FVKHGSRFSSTDKNTEVILG-TVVIAKNPCLHPGDVRILEAVDVPELHHLVDCLVFPQKG 493

Qy        996 DVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLE---PDLSRYLKKDKTTFKQLMA 1052

Db        494 ERPHANEASGSDLDGDLYFVTWDEKLIPPGKKSWNPMDYSPPEA-----------KQL-- 540

Qy       1053 SHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNSVSNKPAIILSSLVG 1112

Db        541 -------PRQVSQHDIIDFFLKNMISENLGRICNAHVVHADLSEYGAMDEKCIHLAELAA 593

Qy       1113 NLVDQSKQGIVFNEASWAQLRRELLGGALS---LPDPM-------YKSDSWLGR------ 1156

Db        594 TAVDFPKTG-----------KLAIMPPHLKPKVYPDFMGKEDGQSYKSEKILGRLYRSIQ 642

Qy       1157 ----GEPTHIIDYLKFSIARP-----AIDKELEAFHNAMKAAKDTEDGAHFWDPDLASYY 1207

Db        643 EASNG------DVVSQEVCTPNDLPYDIDLEVPGASDFLASAWQCKCS---YDAQLSALL 693

Qy       1208 TFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKN--KEMRDSKDPYPVRVNQVYEKW 1265

Db        694 SQY-----RVRTEAELVT------GHIT----FLVKNSSKKQGDIKD----RLKTAY--- 731

Qy       1266 CAITPEAMDKS-GANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHKSPKFVWQM 1324

Db        732 -----SALRKEFKSTFESIASDQCE---IGDDEKNLLYEMKASAWYQVTYH--PKWVEKS 781

Qy       1325 AGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLE 1371

Db        782 RG---------ILGPDGEEIPASLS---FAWIPVD------YLARIK 810
```

```
RESULT 4
AAU10006
ID   AAU10006 standard; protein; 1114 AA.
XX
AC   AAU10006;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Tomato RNA-directed RNA polymerase (RdRP) protein.
XX
KW   Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW   cytostatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW   transgenic plant; transgenic animal; cancer; viral infection;
KW   immunoprecipitation; immunolocalisation; EC 2.7.7.48; gene therapy.
XX
OS   Lycopersicon esculentum.
XX
PN   US2001023067-A1.
XX
PD   20-SEP-2001.
XX
PF   08-FEB-2001; 2001US-00782874.
XX
PR   05-MAR-1997;   97US-00811583.
XX
PA   (WASS/) WASSENEGGER M.
PA   (RIED/) RIEDEL L.
XX
PI   Wassenegger M,  Riedel L,  Schiebel W,  Sanger HL;
XX
DR   WPI; 2001-595798/67.
DR   N-PSDB; AAS17837.
XX
PT   New nucleic acid molecule encoding a polypeptide having the enzymatic
PT   activity of RNA-directed RNA polymerase, for modulating gene expression
PT   and treating cancer and virus infection in human and animals.
XX
PS   Claim 1; Page 19-22; 34pp; English.
XX
CC   This sequence represents the tomato RNA-directed RNA polymerase (RdRP)
CC   protein of the invention. The invention comprises the nucleic acid and
CC   protein sequences of RdRP. The protein of the invention can catalyse in
CC   vitro transcription of short single stranded RNAs into DNA molecules,
CC   this transcription can be either primed by RNA or DNA oligonucleotides or
CC   be unprimed. The protein may have cytostatic or virucide activities. The
CC   sequences of the invention may be used in gene therapy or as an RNA
CC   directed RNA synthesis inhibitor. The RdRP cDNA sequence and a template
CC   nucleic acid molecule derived from a nucleic acid molecule which causes a
CC   disease are useful for treating a disease caused by the undesired
CC   expression or overexpression of a nucleic acid molecule in a human, rat
CC   or mouse, by administering the molecules. This system can be used in the
CC   preparation of a pharmaceutical composition and for inhibiting expression
CC   of any desired gene by transferring the RdRP system to organisms that
CC   either lack a comparable mechanism or do not sufficiently express their
CC   own RdRP. An antibody or an antagonist or inhibitor to the protein are
CC   useful for inhibiting RNA directed RNA synthesis and for ensuring stable
CC   heterologous, gene expression in transgenic organisms. The sequence is
CC   useful for probes and/or for the control of gene expression, as primers
CC   for amplification of nucleic acid molecules and as tools for the
CC   detection of expression of the cDNA molecules. Additionally, nucleotide
CC   and protein sequences are useful for suppression of undesired gene
CC   expression in humans and animals. The RdRP is useful as a therapeutic
CC   agent for the control of cancer and virus infection in humans and animals
CC   and the antibody is useful for immunoprecipitation or immunolocalisation
CC   of the protein, identification of polypeptides interacting with it and
CC   screening expression libraries
XX
SQ   Sequence 1114 AA;

  Query Match             4.0%;  Score 297;  DB 4;  Length 1114;
  Best Local Similarity   20.4%;  Pred. No. 1.2e-14;
  Matches  175;  Conservative  133;  Mismatches  323;  Indels  228;  Gaps   32;

Qy        562 EEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGIT 621

Db        386 EDIDNFLRV---SFVDEEWEKLYSTDL--------------LPK----------ASTGSG 418

Qy        622 FRPDVFK------TRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTW----QPHLKL--- 668

Db        419 VRTNIYERILSTLRKGFVIGDKKFE-FLAFSSSQL------RDNSVWMFASRPGLTANDI 471

Qy        669 ------FSRI---------QLGLSKTYAIMTLEPHQIRHHKTDLL---SPSGTGEVMNDGV 711

Db        472 RAWMGDFSQIKNVAKYAARLGQSFGSSRETL---SVLRHEIEVIPDVKVHGTSYVFSDGI 528

Qy        712 GRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECD 771

Db        529 GKISGDFAHRVASKCGLQYTPSAFQIRYGGYKGVVGVDPDSSMK---LSLRKSMSKYESD 585

Qy        772 FVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQ 831

Db        586 NI-----KLDVLG-WSKYQPCYLNRQLITLLS---------TLGVKDEVLEQKQKEAVDQ 630

Qy        832 KHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYL 891

Db        631 LDAILHDSLKAQEALELMS-----------------PGENTNILKAMLNCGYKPDAEPFL 673

Qy        892 QDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFS----SKFRDEE 947

Db        674 SMMLQTFRASKLLDLRTRSRIFIPNGRTMMGCLDESRTLEYGQVFVQFTGAGHGEFSDDL 733

Qy        948 ESF--------TLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPL 999
```

```
Db        734 HPFNNSRSTNSNFILKGNVVVAKNPCLHPGDIRVLKAVNVRALHHMVDCVVFPQKGKRPH 793

Qy       1000 AKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSA 1059
                : || | |||: :|||| :::          |   |   |   |   |
Db        794 PNECSGSDLDGDIYFVCWDQDMIPPRQVQPMEYPPAPSIQLDHD--------------- 837

Qy       1060 AKEQTTYDMIQKSFHFALQPNFLGMCTN----YKERLCYINNSVSNKPAIILSSLVGNLV 1115
                  | : :::|    :   : ||:   : :|      :     |    |:  |      |
Db        838 ----VTIEEVEEYFTNYIVNDSLGIIANAHVVFADR---EPDMAMSDPCKKLAELFSIAV 890

Qy       1116 DQSKQGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAI 1175
               |   | |:      |    |||  :         ||  |  |  |        : ::|  : :    :  |
Db        891 DFPKTGVPAEIPS--QLRPK------EYPDFMDKPD------KTSYISERVIGKLFRKVK 936

Qy       1176 DKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEV- 1234
              ||                      |   :|  || :::: :|   :  :      :: | |
Db        937 DKA---------------------PQASSIATFTRDVARRSYDADMEVDGFEDYIDEAF 974

Qy       1235 --EKEYGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDKSGANYD--------SKV 1284
                : ||      : :: :    | | ::                |      : |:    :|             |
Db        975 DYKTEY-----DNKLGNLMDYYGIKTE------AEILSGGIMKASKTFDRRKDAEAISVA 1023

Qy       1285 IRLLELSFLADREMNTW---------ALLRASTAFKLYYHKS-------------PKFV 1321
               :| |           :|     |          | :||   :  :  || :                    |
Db       1024 VRALR------KEARAWFKRRNDIDDMLPKASAWYHVTYHPTYWGCYNQGLKRAHFISFP 1077

Qy       1322 WQMAGRQLAYIKAQMTSRP 1340
               | :   : :      | :   :||
Db       1078 WCVYDQLIQIKKDKARNRP 1096
```

RESULT 5
AAE00897
ID   AAE00897 standard; protein; 1114 AA.
XX
AC   AAE00897;
XX
DT   04-JUL-2001 (first entry)
XX
DE   Tomato C-protein having RNA-directed RNA polymerase (RdRP) activity.
XX
KW   Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW   transgenic plant; tissue culture; plant breeding; therapy; C-protein.
XX
OS   Lycopersicon esculentum.
XX
FH   Key   Location/Qualifiers
FT   Region   331. .346
FT   /label= P432
FT   /note= "C-protein-specific antibody for immunisation of
FT   rabbits"
FT   Region   377. .390
FT   /label= P430
FT   /note= "C-protein-specific antibody for immunisation of
FT   rabbits"
FT   Region   784. .797
FT   /label= P433
FT   /note= "C-protein-specific antibody for immunisation of
FT   rabbits"
FT   Region   1007. .1020
FT   /label= P431
FT   /note= "C-protein-specific antibody for immunisation of
FT   rabbits"
XX
PN   US6218142-B1.
XX
PD   17-APR-2001.
XX
PF   05-MAR-1997;   97US-00811583.
XX
PR   05-MAR-1997;   97US-00811583.
XX
PA   (WASS/) WASSENEGGER M.
PA   (RIED/) RIEDEL L.
XX
PI   Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
DR   WPI; 2001-289830/30.
DR   N-PSDB; AAD04370.
XX
PT   New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT   polymerase enzymatic activity, useful in modulating gene expression in
PT   plants, humans and animals, as well as in plant cell/tissue cultures or
PT   plant breeding.
XX
PS   Claim 1; Col 33-40; 31pp; English.
XX
CC   The present sequence is a tomato C-protein having RNA-directed RNA
CC   polymerase (RdRP) activity. This protein is capable of RNA-directed RNA
CC   synthesis, thus using RNA as a template for synthesising complementary
CC   RNA molecules. RdRP nucleic acid is useful for modulating gene expression
CC   in plants, humans and animals. This may lead to various physiological,
CC   developmental and/or morphological changes. Transgenic plants containing
CC   RdRP nucleic acid is especially useful in plant cell or tissue cultures
CC   and in plant breeding. RdRP is useful in gene therapy, particularly for
CC   treating a disease that is caused by the undesirable expression or
CC   overexpression of a gene
XX
SQ   Sequence 1114 AA;

```
  Query Match             4.0%;  Score 297;  DB 4;  Length 1114;
  Best Local Similarity   20.4%;  Pred. No. 1.2e-14;
  Matches  175;  Conservative  133;  Mismatches  323;  Indels  228;  Gaps   32;

Qy        562 EEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGIT 621
              |::  :| :     |  |  :|    :: |                  ||          | ||
Db        386 EDIDNFLRV---SFVDEEWEKLYSTDL-------------LPK---------ASTGSG 418

Qy        622 FRPDVFK------TRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTW----QPHLKL--- 668
               | ::::      :  |  ::  |    |  ||:       ||: |    :| |
Db        419 VRTNIYERILSTLRKGFVIGDKKFE-FLAFSSSQL------RDNSVWMFASRPGLTANDI 471

Qy        669 ------FSRI--------QLGLSKTYAIMTLEPHQIRHHKTDLL---SPSGTGEVMNDGV 711
                    ||:|        :||  |    :   ||     :   |: :::       ||  |  :||:
Db        472 RAWMGDFSQIKNVAKYAARLGQSFGSSRETL---SVLRHEIEVIPDVKVHGTSYVFSDGI 528

Qy        712 GRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECD 771
              |::|   | |:     ||   |||  | |:   ||: :|  | :      :      |  |:|  |
Db        529 GKISGDFAHRVASKCGLQYTPSAFQIRYGGYKGVVGVDPDSSMK---LSLRKSMSKYESD 585

Qy        772 FVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQ 831
                :      |:|     |: :    || ||: :|                :  | ::    |:: :|
Db        586 NI-----KLDVLG-WSKYQPCYLNRQLITLLS---------TLGVKDEVLEQKQKEAVDQ 630

Qy        832 KHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYL 891
                |:          |   |  |                     |       |   ::| |:  |   : :|
Db        631 LDAILHDSLKAQEALELMS----------------PGENTNILKAMLNCGYKPDAEPFL 673

Qy        892 QDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFS----SKFRDEE 947
                 :    :    |   |:::  | :       :         |     ||  :| | |:        :| |:
Db        674 SMMLQTFRASKLLDLRTRSRIFIPNGRTMMGCLDESRTLEYGQVFVQFTGAGHGEFSDDL 733

Qy        948 ESF--------TLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPL 999
                |              :      :|:||::|    | ||: ::||    || :  |  ::|  || |
Db        734 HPFNNSRSTNSNFILKGNVVVAKNPCLHPGDIRVLKAVNVRALHHMVDCVVFPQKGKRPH 793

Qy       1000 AKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSA 1059
                : || | |||: :|||| :::          |   |   |   |   |
Db        794 PNECSGSDLDGDIYFVCWDQDMIPPRQVQPMEYPPAPSIQLDHD--------------- 837

Qy       1060 AKEQTTYDMIQKSFHFALQPNFLGMCTN----YKERLCYINNSVSNKPAIILSSLVGNLV 1115
                  | : :::|    :   : ||:   : :|      :     |    |:  |      |
Db        838 ----VTIEEVEEYFTNYIVNDSLGIIANAHVVFADR---EPDMAMSDPCKKLAELFSIAV 890

Qy       1116 DQSKQGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAI 1175
               |   | |:      |    |||  :         ||  |  |  |        : ::|  : :    :  |
```

```
Db        891 DFPKTGVPAEIPS--QLRPK------EYPDFMDKPD------KTSYISERVIGKLFRKVK 936

Qy       1176 DKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEV- 1234

Db        937 DKA---------------------PQASSIATFTRDVARRSYDADMEVDGFEDYIDEAF 974

Qy       1235 --EKEYGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDKSGANYD--------SKV 1284

Db        975 DYKTEY-----DNKLGNLMDYYGIKTE------AEILSGGIMKASKTFDRRKDAEAISVA 1023

Qy       1285 IRLLELSFLADREMNTW---------ALLRASTAFKLYYHKS--------------PKFV 1321

Db       1024 VRALR------KEARAWFKRRNDIDDMLPKASAWYHVTYHPTYWGCYNQGLKRAHFISFP 1077

Qy       1322 WQMAGRQLAYIKAQMTSRP 1340

Db       1078 WCVYDQLIQIKKDKARNRP 1096
```

RESULT 6
AAB28531
ID AAB28531 standard; protein; 1164 AA.
XX
AC AAB28531;
XX
DT 07-FEB-2001 (first entry)
XX
DE Maize RNA-directed RNA polymerase, SEQ ID NO: 8.
XX
KW Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW plant virus resistance; plant breeding.
XX
OS Zea mays.
XX
PN WO200060097-A1.
XX
PD 12-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US009105.
XX
PR 07-APR-1999; 99US-0128094P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Odell JT, Orozco EM, Weng Z, Sakai H, Cahoon RE, Wang J;
PI Zhong G;
XX
DR WPI; 2000-679376/66.
DR N-PSDB; AAC63740.
XX
PT New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT for controlling gene expression and providing mechanisms to engineer
PT plant virus resistance.
XX
PS Claim 10; Page 41-45; 62pp; English.
XX
CC The present sequence is part of a plant RNA-directed RNA polymerase
CC protein. Polynucleotides encoding RNA-directed RNA polymerases were
CC isolated from plant cDNA libraries. They are useful as probes for
CC genetically and physically mapping genes, and as markers for traits
CC linked to those genes. They are useful for controlling gene expression
CC and provide mechanisms to engineer plant virus resistance. They are also
CC useful for plant breeding to develop lines with desired phenotypes
XX
SQ Sequence 1164 AA;

```
  Query Match             3.9%;  Score 292;  DB 3;  Length 1164;
  Best Local Similarity   21.4%;  Pred. No. 3.5e-14;
  Matches  176;  Conservative  123;  Mismatches  323;  Indels  200;  Gaps   35;

Qy        605 KPIIKERVHFFAETG----------ITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWL 654

Db        453 KTTVYKRVKTFLTEGFHMCGRKYSFLAFSSNQLRDRS---AWFFAEDRTT-TVETIRKWM 508

Qy        655 LQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRM 714

Db        509 GRFTSKNVAKHA---ARMGQCFSSTYATVVLQPHEVNECLDEV---EHNGYIFSDGIGKI 562

Qy        715 SRSVAKRIRDVLGLGD-VPSAVQGRFGSAKGMWVIDV-DDTGDEDWIETYPSQRKWECDF 772

Db        563 TCDLALEVAQKLQLTDNPPSAYQIRYAGFKG--VISVWEGKNDGIRLSLRPSMHKFE--- 617

Qy        773 VDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQK 832

Db        618 --SNHTVLEVVS-WTKFQPGFLNRQIITLLS---------------------------- 645

Qy        833 HALNRPVEFRQWVYESYSSRATRV-SHGRVPF--LAGLPDSQEETLNFLMNSGFDPKKQK 889

Db        646 -SLNVPDAIFAQMQEAMLSNLNNILSDSDVAFDIVTASCAEQGTTAALMLSAGISPGTEP 704

Qy        890 YLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEES 949

Db        705 HLKAMLLAIRSSQLLGLLEKTRIFVPKGRWLMGCLDELGILEQGQCFIRASS-------- 756

Qy        950 FTLLSDC--------------------DVLVARSPAHFPSDIQRVRAVFKPELHSLKDVI 989

Db        757 -PSLNNCLVKYGSRLSAANTNAETILGTIVMAKNPCLHPGDVRILEAVDVPELHHLVDCL 815

Qy        990 IFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQ 1049

Db        816 VFPKKGERPHANEASGSDLDGDLYFVTWDENLIPPGKKSWNPMD-----YSPAEAKQLPR 870

Qy       1050 LMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNS----------- 1098

Db        871 AVSQH-----------DIV--GF-------FLKNMVN--EKLGPISNAHVVHADMSEYG 907

Qy       1099 VSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDPM-YKSDSWLGR- 1156

Db        908 AMDEKCIQLAELAATAVDFPKTGKIVSMP--ASLRPKLYPDFMGKEDAISYRSEKILGRL 965

Qy       1157 ---------------------------GEPTHIIDYLKFS-IARPAIDKELEAFHNAMK 1187

Db        966 YRSIQEASSDDLVPEETCTSNNLPYDADMEVAGAADFLSSAWQCKCSYETQLNALLNQYG 1025

Qy       1188 AAKDTE-DGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKE 1246

Db       1026 VRTEAELVTEHIW--SLPKY---------SSRRQGDIKERLKNAYYALHKEF-RSIFESI 1073

Qy       1247 MRDSKDPYPVRVNQVYE----KWCAIT--PEAMDKSGANYDSKVIRL-LELSFLADREMN 1299

Db       1074 VTDQTEISDDEKSRFYEMKASAWYQVTYHPEWVQKSREMFKSDCEDMPARLSF------- 1126

Qy       1300 TWALLRASTAFKLYYH-----KSPKFVWQMAGRQLAYIKAQM 1336

Db       1127 AWIAVEHLARIKIRCRGEVKVDSPRPVERLA----AYISGSM 1164
```

RESULT 7
AAB28533
ID AAB28533 standard; protein; 1125 AA.
XX
AC AAB28533;
XX
DT 07-FEB-2001 (first entry)
XX
DE Soybean RNA-directed RNA polymerase, SEQ ID NO: 12.
XX
KW Soybean; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW plant virus resistance; plant breeding.
XX
OS Glycine max.
XX
PN WO200060097-A1.
XX
PD 12-OCT-2000.

```
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT,  Orozco EM,  Weng Z,  Sakai H,  Cahoon RE,  Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   N-PSDB; AAC63742.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 10; Page 50-54; 62pp; English.
XX
CC   The present sequence is part of a plant RNA-directed RNA polymerase
CC   protein. Polynucleotides encoding RNA-directed RNA polymerases were
CC   isolated from plant cDNA libraries. They are useful as probes for
CC   genetically and physically mapping genes, and as markers for traits
CC   linked to those genes. They are useful for controlling gene expression
CC   and provide mechanisms to engineer plant virus resistance. They are also
CC   useful for plant breeding to develop lines with desired phenotypes
XX
SQ   Sequence 1125 AA;

  Query Match            3.8%;  Score 280;  DB 3;  Length 1125;
  Best Local Similarity  19.0%;  Pred. No. 3.3e-13;
  Matches  207;  Conservative  144;  Mismatches  355;  Indels  386;  Gaps   44;

Qy        344 SSQGETSFSTYYESFPSSG---------GEGAIPEP----------------SRSNGLAR 378
Db         12 TSVNVSDVKTFVEQYTGEGTVFAIKLRHGKGRVPRAFAIIQFTTANSATSMMSRANNILR 71

Qy        379 SEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGL 438
Db         72 TLRYGTSYLKAR-----EMERDIVPR----------------PRVFLHSLDDVK-----L 105

Qy        439 KYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTG----NFESKGSAVVL 494
Db        106 SFGCQISKGR-FSVLWK-----------------KQDVIVNFGSGMRKMHFLFSHNNVQY 147

Qy        495 SAVLDY---------NPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTS 545
Db        148 KLELSYENIWKIELHRPRNETTR--YL-------------LIQLLGAPRVFENDV--PTS 190

Qy        546 TSPSVPPVVSKQPGAVEEVIQWL---------TMGQHSLV------GRQWRAFFAKDAGY 590
Db        191 TNIFDDPLFNFFKDAPDE--QWIRAIDFTPESRIGQSSAICLELPNGRQLPNFRENFAYY 248

Qy        591 RKPLREFQL--------------------------------------------------- 599
Db        249 EESERQYTLQTGVPFSQNWGLVPIVAPPLGVKISYDILFKVNSLVQHACLAGPALDGDFY 308

Qy        600 RAEDPKPIIKERVHFFAE-----TGITFRP-----DVFKT--------RS---------- 631
Db        309 RLVDPRRMPREFIEYALEKIYYSKEFCYEPTKWLTDQYKTYLESKNHPRSPAISLDTGLV 368

Qy        632 ------VVPAE-------------------EPVEQ--RTEFKVSQMLDWLLQLD------ 658
Db        369 YVRRVQITPCKVYFCGPEMNVSNRVLRHFREHIDNFLRVSF-VDEELDKLFSTDLSSRSQ 427

Qy        659 NNTWQPHLKLFSRIQLGL---SKTYAIMTLEPHQIRHHKTDLLSPSGTG----------- 704
Db        428 NKKTEIYTRILSILKNGIVVGDKKFEFLAFSSSQLRENSLWMFAPTETGCTAAYIRKWMG 487

Qy        705 -----------------------------------------------EVMNDGVGRMSR 716
Db        488 NFSQIRNVAKYAARLGQSFGSSTETLSVHRDEVEIIPDVKKLTYDGNEYVFSDGIGKISL 547

Qy        717 SVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKH 776
Db        548 EFAQKVAKKCGYDCTPSAFQIRYGGYKGVVAVDPKSCYK---LSLRKSMRKYDSD----- 599

Qy        777 QRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALN 836
Db        600 NTKLDVLA-RSKFQPCYLNRQLISLLS---------TLGIKDDVFEKKQRETVNQLNTIL 649

Qy        837 RPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAW 896
Db        650 TDSLKAQEVLDLMSA-------GEI----------TNVLKEMLICGYKPNEEPFLSMMLQ 692

Qy        897 DLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSS----KFRDEEESFTL 952
Db        693 TFRASKLLELRLKSRIFIPKGRAMMGCLDETRTLEYGQVFVQFSNNRLQNLSDDFFSYDL 752

Qy        953 ----LSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDY 1008
Db        753 PKNYMVKGKVVVAKNPCLHPGDVRVLQAVDVPDLYHMVDCVVFPQKGPRPHPNECSGSDL 812

Qy       1009 DGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDM 1068
Db        813 DGDIYFVCWDHELI-----PSRPIDP-------MDYTAPATVELDH--------DVMIEE 852

Qy       1069 IQKSFHFALQPNFLGMCTNY------KERLCYINNSVSNKPAIILSSLVGNLVDQSKQGI 1122
Db        853 VEEYFANYIVNDSLGIIANAHTVFADKEHL-----KAMSDQCVKLARLFSTAVDFPKTGV 907

Qy       1123 VFNEASWAQLRREL----LGGALSLPD-PMYKSDSWLGR---------GEPTHIIDYLKF 1168
Db        908 P------AVIPPELHVKEYPDFMEKPDKPTYKSHNVIGKLFREVKEISTSAGSITSFTKL 961

Qy       1169 SIARPAIDKELE 1180
Db        962 -VARDSYDHEME 972

RESULT 8
AAB28530
ID   AAB28530 standard; protein; 1108 AA.
XX
AC   AAB28530;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Maize RNA-directed RNA polymerase, SEQ ID NO: 6.
XX
KW   Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW   plant virus resistance; plant breeding.
XX
OS   Zea mays.
XX
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT,  Orozco EM,  Weng Z,  Sakai H,  Cahoon RE,  Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   N-PSDB; AAC63739.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
```

```
XX
PS   Claim 10; Page 36-40; 62pp; English.
XX
CC   The present sequence is part of a plant RNA-directed RNA polymerase
CC   protein. Polynucleotides encoding RNA-directed RNA polymerases were
CC   isolated from plant cDNA libraries. They are useful as probes for
CC   genetically and physically mapping genes, and as markers for traits
CC   linked to those genes. They are useful for controlling gene expression
CC   and provide mechanisms to engineer plant virus resistance. They are also
CC   useful for plant breeding to develop lines with desired phenotypes
XX
SQ   Sequence 1108 AA;

  Query Match             3.6%;  Score 267.5;  DB 3;  Length 1108;
  Best Local Similarity   21.7%;  Pred. No. 3.5e-12;
  Matches  155;  Conservative  113;  Mismatches  304;  Indels  141;  Gaps   26;

Qy        645 FKVSQMLD---W-------LLQLDNNTWQPHLKLFSRIQ-----LGLSKTYAIMTLEPHQ 689
Db        465 FSSSQLRDNSAWMFASRQGLTASDIRKWMGDFRDIRNVAKYAAILGQSFSSSTETLKVHK 524

Qy        690 IRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGD-VPSAVQGRFGSAKGMWVI 748
Db        525 SEVERIPDIT-NGTKYIFSDGVGKISANFAVEVAMKCKLKRFAPSVFQIRYGGYKGVVAV 583

Qy        749 DVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARD 808
Db        584 ---DTRSNHKLSLRKSMSKFQSENI-----TLDVLAY-SKYQPCFLNRQLITLLS----- 629

Qy        809 KVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLP 868
Db        630 ----TLGVSDNVFELKQKEALRQ---LNRMVT---------EPQAAREAVELMPM----- 668

Qy        869 DSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWG 928
Db        669 GEVTNVVKELLSCGYQPDHEPYLSMLLQTFRASKLLELKTKSRIFITQGRAMMGCLDETC 728

Qy        929 VLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDV 988
Db        729 TLKYGQVFVQASYSADDHRK---VVVTGKVVVAKNPCLHPGDIRVLQAVDVPALHHLFDC 785

Qy        989 IIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFK 1048
Db        786 VVFPQQGPRPHPNECSGSDLDGDIYFVSWDPHLIPSRLVDPMDYTPASAETLDHD----- 840

Qy       1049 QLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNY------KERLCYINNSVSNK 1102
Db        841 ---------------VTIEEIQEYFTNYIVNESLGIIANAHVVFTDQERM-----KAESP 880

Qy       1103 PAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDPM-------YKSDSWLG 1155
Db        881 PCVQLAKLFSIAVDFPKTGVP------ALIPHEL--HVKEYPDFMEKLDKVTYESKGVIG 932

Qy       1156 R-----GEPTHIIDYLKFSIARPAIDKEL--EAFHNAMKAAKDTEDGAHFWDPDLASYYT 1208
Db        933 KLYREIKKHTPHIKHFTREVARRSYDTDLIVDGYEDYITEAIEFKEEYDFRLGNLMDHYG 992

Qy       1209 FFKE---ISDKSRSSALLFT------TLKNRIGEVEKE----YGRLVKNKEMRDSKDP-- 1253
Db        993 IKSEAEIISGCILKMAKNFTKSSDADAIRMAVRSLRKEARSWFNEMSTGEDGQDAMEAKA 1052

Qy       1254 ---YPVRVNQVY---------------EKWCAITPEAMDKSGANYDSKVIRLL 1288
Db       1053 SAWYHVTYHQQYWGSYNEGYDRPHLISFPWCVYDKLVAIKQGRNLLTQMDRNL 1105

RESULT 9
AAB28529
ID   AAB28529 standard; protein; 417 AA.
XX
AC   AAB28529;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Maize partial RNA-directed RNA polymerase, SEQ ID NO: 4.
XX
KW   Maize; plant; RNA-directed RNA polymerase; gene mapping; gene marker;
KW   plant virus resistance; plant breeding.
XX
OS   Zea mays.
XX
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT,   Orozco EM,   Weng Z,   Sakai H,   Cahoon RE,   Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   N-PSDB; AAC63738.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 10; Page 34-35; 62pp; English.
XX
CC   The present sequence is part of a plant RNA-directed RNA polymerase
CC   protein. Polynucleotides encoding RNA-directed RNA polymerases were
CC   isolated from plant cDNA libraries. They are useful as probes for
CC   genetically and physically mapping genes, and as markers for traits
CC   linked to those genes. They are useful for controlling gene expression
CC   and provide mechanisms to engineer plant virus resistance. They are also
CC   useful for plant breeding to develop lines with desired phenotypes
XX
SQ   Sequence 417 AA;

  Query Match             3.2%;  Score 236;  DB 3;  Length 417;
  Best Local Similarity   21.3%;  Pred. No. 2.7e-10;
  Matches   84;  Conservative   74;  Mismatches  158;  Indels   78;  Gaps   11;

Qy        698 LSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDV--PSAVQGRFGSAKGMWVIDVDDTGD 755
Db         26 VTTDGTKYIFSDGIGKISTRFARQVAKLIGLDPAHPPSAFQIRYGGYKG--VITIDPTSF 83

Qy        756 EDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNL----QLLPVLEDRARDKVK 811
Db         84 FN-LSLRPSMKKFES-------------------KSTMLNITNWSKSQPCYVNREIISLL 123

Qy        812 MRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQ 871
Db        124 STLGIKDEVFXSMQQDDMHESDGMLTNKE-----------------AALSVLGKIGGGD 165

Qy        872 EETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLE 931
Db        166 TKTAADMLLQGYEPSSEPYLLMILKAHRANRLTDIRTRCKIHVQKGRVLIGCLDETCKLE 225

Qy        932 ENEVHVGFSSKFRDE---EESF-------TLLSDCDVLVARSPAHFPSDIQRVRAVFKPE 981
Db        226 YGQVYIRITKNRKEQKYSEQPFFCNDDGKTAVIVGKVAITKNPCLHPGDVRVLEAVYDPG 285

Qy        982 L--HSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRY 1039
Db        286 LDARGLIDCVVFPQRGERPHPNECSGGDLDGDLFFITWDDKLIPEKVDAPM--------- 336

Qy       1040 LKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSF 1073
Db        337 ---DYTATRPRIMDHA--------VTLEEIQKHF 359
```

```
RESULT 10
AAY78795
ID   AAY78795 standard; protein; 1219 AA.
XX
AC   AAY78795;
XX
DT   19-MAY-2000  (first entry)
XX
DE   Human antizuai-2 (AZ-2) amino acid sequence.
XX
KW   Antizuai-2; AZ-2; human; breast cancer; malignancy progression marker;
KW   malignancy reversion; tumour suppressor.
XX
OS   Homo sapiens.
XX
PN   WO200000503-A1.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US014482.
XX
PR   26-JUN-1998;   98US-0090747P.
XX
PA   (CHEN/) CHEN H.
PA   (BISS/) BISSELL M.
XX
PI   Chen H,  Bissell M;
XX
DR   WPI; 2000-170903/15.
DR   N-PSDB; AAZ90112.
XX
PT   New nucleic acid encoding a tumor suppressor or marker, used for
PT   diagnosis, monitoring progress or treatment, and gene therapy of breast
PT   cancer.
XX
PS   Claim 6; Page 69-72; 120pp; English.
XX
CC   This sequence represents the human antizuai-2 (AZ-2) amino acid sequence.
CC   The AZ-2 gene is a variant of the antizuai-1 (AZ-1) gene (see AAZ90111).
CC   The AZ-1 gene is located on chromosome 10q26, and encodes a protein that
CC   acts as a tumour suppressor or marker of malignancy progression or
CC   reversion. The AZ-1 protein and its variants are tumour suppressors, AZ-1
CC   interacts with E-cadherin and beta-catenin. Detecting low levels of AZ-1
CC   nucleotide or amino acid sequences are used to diagnose a breast cell
CC   malignancy, also for monitoring disease progression, particularly
CC   assessment of therapeutic efficacy. The nucleotide sequence is used in in
CC   vivo or ex vivo gene therapy, and AZ-1 polypeptides are used for treating
CC   or preventing breast cancer. AZ-1 polypeptides are also used to raise
CC   specific antibodies, for diagnostic detection of AZ-1. Fragments of the
CC   AZ-1 nucleotide sequence are useful as probes or primers for detecting
CC   expression of the AZ-1 gene
XX
SQ   Sequence 1219 AA;

  Query Match             2.5%;  Score 182;  DB 3;  Length 1219;
  Best Local Similarity   18.8%;  Pred. No. 5.7e-05;
  Matches 195;  Conservative  155;  Mismatches  382;  Indels  306;  Gaps   47;

Qy          1 MNPITPRKRNSPVEEIINRLNNDYNLGLQ-----CVADTTLTPHRRKELAESDEDFGRHD 55
                | |  |  | :::|   ::::: :|    | :      |||  :  |    :|
Db        226 MLPSVP-KKDAP--RVMDKVTSDETRGAEGTESSPVADDIIQPAAPADL---------ES 273

Qy         56 KIYRALNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLPWSKEPPRAATAGQ 115
                     | ::        | : |    :    :   :  |        :|    |::    | ||
Db        274 PTLAASSY----HSDVVGQVSTDLIAQRSSDSE----EAFETPESTTPVKAPPAPP---- 321

Qy        116 QWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNTKRKDEPA----NVTF 171
                                    |||                 |  :  |   :   :||      |
Db        322 ----------------PPP-------------PEVIPEPEVSTQPPPEEPGCGSETVPV 351

Qy        172 ADPPKR-SLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLN-----------QC 218
                 | |:  |:  |    || |  :        |:    |     :|: :|:

Db        352 PDGPRSDSVEGSPFRPPSHPFSAVFDEDQPI-ASSGTYNLDFDNIELVDTFQTLEPRASD 410

Qy        219 TKRAKGKLSDNVAAAAAPPVPIA-SALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQG 277
                | :||:  |     :    |||: | |  :  :  : ::       :::|:  |   :      | |
Db        411 AKNQEGKV--NTRRKSTDSVPISKSTLSRSLSLQASDFDGASSSGNPEAVALAPDAYSTG 468

Qy        278 TSYGSSVFSACR-------HNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGR 330
                :|  ||       :             :  ||:    | ||  :  :: |  : :: :|  :  | :|
Db        469 SSSASSTLKRTKKPRPPSLKKKQTTKKPTETPPVKETQQEPDEESLVPSGENLASETKTE 528

Qy        331 TTKSH------IDNQPLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESAR 384
                : |:         ::  ||                        |: |  :  |  |  |  |        |
Db        529 SAKTEGPSPALLEETPLE----------------PAVGPKAACPLDSESAEGVVPPASGG 572

Qy        385 SQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDE-----SLGLK 439
                :||    ||    |                   ||  |  |    :          ||        |: |:
Db        573 GRVQNSPPVGRKTL--------------PLTTAPEAGEVTPSDSGGQEDSPAKGLSVRLE 618

Qy        440 YD-----PSWSTARD------------------VTDIWKTLYRLDAFRGKPFPEKP--P 473
                :|         ||    ::                     :  ||   :||        |  |  |   |
Db        619 FDYSEDKSSWDNQQENPPPTKKIGKKPVAKMPLRRPKMKKTPEKLDNTPASP-PRSPAEP 677

Qy        474 NDVFVTAMTGNFESKGSAVVLSAVLDY--NPDNSPTAPLYLVKLKPLMFEQGCRLTRRFG 531
                ||: :      |   |:               :|      :|: :  :|  :         ::    |  : :|          :   |
Db        678 NDIPIAKGTYTFD-----------IDKWDDPNFNPFSSTSKMQESPKLPQQ----SYNFD 722

Qy        532 PDRFFEILIPSPTST-SPSVPPVVSKQPGAVEEVIQWLTMGQHSLVG---------RQWR 581
                ||    |  :  |    ||: :||  |    ||  |  :  |     |        |  : :  |              ::
Db        723 PDTCDESVDPFKTSSKTPSSP---SKSPASFE--IPASAMEANGVDGDGLNKPAKKKKTP 777

Qy        582 AFFAKDAGYRKPLREFQLRAEDP--------KPIIKERVHFFAETGI------------ 620
                  |  :     |  ||   :       ::||         |:|          ||       |        :
Db        778 LKTVKKSPKRSPLSD--PPSQDPTPAATPETPPVISAVVHATDEEKLAVTNQKWTCMTVD 835

Qy        621 ---------------TFRPDVFKTRSVVPAEEPVEQRTEFKVSQM-------------L 651
                                    ||     |  :|:     |  ||   ::  |     :::    |
Db        836 LEADKQDYPQPSDLSTF---VNETKFSSPTEE-LDYRNSYEIEYMEKIGSSLPQDDDAPK 891

Qy        652 DWLLQLDNNTWQPHLKLFSRIQLGLS----------KTYAIMT-----------LEPHQ 689
                  | |    :|  |           | :::    |                         :|  |:::                       | |:|
Db        892 KQALYLMFDTSQESPVKSSPVRMSESPTPCSGSSFEETEALVNTAAKNQHPVPRGLAPNQ 951

Qy        690 IRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDV---PSAVQGRFGSAK--- 743
                |     |  |  |     :   :   :   :|   |     |  |   ||:        ||      :|:    |  |:|:
Db        952 ESH----LQVPEKSSQKELEAMGLGTPSEAIEIREAAHPTDVSISKTALYSRIGTAEVEK 1007

Qy        744 --GMWV--IDVDDTGDEDWIETYPSQR---KWECDFVDKHQRTLEVRSVASELKSAGLNL 796
                       |:         |             :|:   :      :  :   :  :   :|:|  : :|  :
Db       1008 PAGLLFQQPDLDSALQIARAEIITKEREVSEWKDKYEESRREVMEMRKIVAEYEKT---- 1063

Qy        797 QLLPVLEDRARDK-----------VKMRQAIGD-----RLINDLQRQFSEQKHAL---NR 837
                  :   ::||    |:|                     ::    ||:  |        :  :  ||  |::  :   |    |          :
Db       1064 -IAQMIEDEQREKSVSHQTVQQLVLEKEQALADLNSVEKSLADLFRRYEKMKEVLEGFRK 1122

Qy        838 PVEFRQWVYESYSSRATR 855
                  |   :      :  |  ||  :
Db       1123 NEEVLKRCAQEYLSRVKK 1140


RESULT 11
AAM38679
ID   AAM38679 standard; protein; 966 AA.
XX
AC   AAM38679;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 1824.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
```

```
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US034263.
XX
PR   23-DEC-1999;   99US-00471275.
PR   21-JAN-2000; 2000US-00488725.
PR   25-APR-2000; 2000US-00552317.
PR   20-JUN-2000; 2000US-00598042.
PR   19-JUL-2000; 2000US-00620312.
PR   03-AUG-2000; 2000US-00653450.
PR   14-SEP-2000; 2000US-00662191.
PR   19-OCT-2000; 2000US-00693036.
PR   29-NOV-2000; 2000US-00727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J,  Zhao QA;
PI   Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI57835.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders such
PT   as central nervous system injuries.
XX
PS   Example 3; SEQ ID NO 1824; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC   encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders. Note: The sequence data for this patent did not form
CC   part of the printed specification
XX
SQ   Sequence 966 AA;

  Query Match             2.4%;  Score 179.5;  DB 4;  Length 966;
  Best Local Similarity   19.3%;  Pred. No. 6.2e-05;
  Matches  175;  Conservative  133;  Mismatches  333;  Indels  265;  Gaps   40;

Qy        132 PPPNNTPGRTFGRTLSGPSGLSRPTSTNTKRKDEPA----NVTFADPPKR-SLTRSATGP 186
              |||            |  :  |  :     :||      |   | |:  |:  |   |
Db         65 PPP-------------PEVIPEPEVSTQPPPEEPGCGSETVPVPDGPRSDSVEGSPFRP 110

Qy        187 PIHGAAIPLKFPDPVNTGSKRPSLESENLN-----------QCTKRAKGKLSDNVAAAA 234
              | |  :      |:  |    :|: :|:                |  :||:  |    :
Db        111 PSHSFSAVFDEDKPI-ASSGTYNLDFDNIELVDTFQTLEPRASDAKNQEGKV--NTRRKS 167

Qy        235 APPVPIA-SALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACR---- 289
                 |||: |  |  :  : : ::    :::|:  |   :     | |:|  ||     :
Db        168 TDSVPISKSTLSRSLSLQASDFDGASSSGNPEAVALAPDAYSTGSSSASSTLKRTKKPRP 227

Qy        290 ---HNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKSH------IDNQP 340
                 : ||:   | ||  :  ::  | : ::  :  : |  :|  : |:       ::  |
Db        228 PSLKKKQTTKKPTETPPVKETQQEPDEESLVPSGENLASETKTESAKTEGPSPALLEETP 287
```

```
Qy        341 LSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRN 400
              |                |: | : |  |   | |          |    :||   ||    |
Db        288 LE---------------PAVGPKAACPLDSESAEGVVPPASGGGRVQNSPPVGRKTL-- 329

Qy        401 IWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDE-----SLGLKYD-----PSWSTARD- 449
                         || |  |    :        ||       |: |::|       ||    ::
Db        330 -----------PLTTAPEAGEVTPSDSGGQEDSPAKGLSVRLEFDYSEDKSSWDNQQEN 377

Qy        450 -----------------VTDIWKTLYRLDAFRGKPFPEKP--PNDVFVTAMTGNFESKG 489
                         :  ||  :||         |   |    | |||: :   |    |:
Db        378 PPPTKKIGKKPVAKMPLRRPKMKKTPEKLDNTPASP-PRSPAEPNDIPIAKGTYTFD--- 433

Qy        490 SAVVLSAVLDY--NPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTST- 546
                    :|   :|: :| :    ::   | : :|    :   | ||    | :  |  ||:
Db        434 --------IDKWDDPNFNPFSSTSKMQESPKLPQQ----SYNFDPDTCDESVDPFKTSSK 481

Qy        547 SPSVPPVVSKQPGAVEEVIQWLTMGQHSLVG------------RQWRAFFAKDAGYRKP 593
              :||  |    || | : |   |      |   : : |            :   |   | :  | |
Db        482 TPSSP---SKSPASFE--IPASAMEANGVDGDGLNKPAKKKKTPLKTDTFRVKKSPKRSP 536

Qy        594 LREFQLRAEDP--------KPIIKERVHFFAETGI----------------------- 620
              | :   ::||          |:|   ||   |   :
Db        537 LSD--PPSQDPTPAATPETPPVISAVVHATDEEKLAVTNQKWTCMTVDLEADKQDYPQPS 594

Qy        621 ---TFRPDVFKTRSVVPAEEPVEQRTEFKVSQM-------------LDWLLQLDNNTWQ 663
                 ||   | :|:   | || ::  |   :::   |                 |  |  :|  |
Db        595 DLSTF---VNETKFSSPTEE-LDYRNSYEIEYMEKIGSSLPQDDDAPKKQALYLMFDTSQ 650

Qy        664 PHLKLFSRIQLGLS----------KTYAIMT------------LEPHQIRHHKTDLLSPS 701
                    | :::  |                :| |::                | |:|  |     |   |
Db        651 ESPVKSSPVRMSESPTPCSGSSFEETEALVNTAAKNQHPVPRGLAPNQESH----LQVPE 706

Qy        702 GTGEVMNDGVGRMSRSVAKRIRDVLGLGDV---PSAVQGRFGSAK-----GMWV--IDVD 751
              : :    : :|    :   |  |   ||:        ||       :|   |  |:|:        |:   :|
Db        707 KSSQKELEAMGLGTPSEAIEIREAAHPTDVSISKTALYSRIGTAEVEKPAGLLFQQPDLD 766

Qy        752 DTGDEDWIETYPSQR---KWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARD 808
                         |     :|  :   :     :|:|  : :|  :             :  ::||  |:
Db        767 SALQIARAEIITKEREVSEWKDKYEESRREVMEMRKIVAEYEKT-----IAQMIEDEQRE 821

Qy        809 K-----------VKMRQAIGD-----RLINDLQRQFSEQKHAL---NRPVEFRQWVYESY 849
              |            ::  ||: |    : :  || |::  :  |    |      :  |  :   : |
Db        822 KSVSHQTVQQLVLEKEQALADLNSVEKSLADLFRRYEKMKEVLEGFRKNEEVLKRCAQEY 881

Qy        850 SSRATR 855
               || :
Db        882 LSRVKK 887
```

```
RESULT 12
ABJ25640
ID   ABJ25640 standard; protein; 1555 AA.
XX
AC   ABJ25640;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene protein #298.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013142.
XX
PR   23-APR-2001; 2001US-0285697P.
PR   27-APR-2001; 2001US-0287066P.
PR   05-JUN-2001; 2001US-0295890P.
```

```
PR   09-JUL-2001; 2001US-0303899P.
PR   31-AUG-2001; 2001US-0316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of Aspergillus
PT   fumigatus, useful for treating or preventing infections by A. fumigatus,
PT   or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS   Disclosure; Page; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This sequence represents a protein of one of the essential genes
CC   of Aspergillus fumigatus of the invention
XX
SQ   Sequence 1555 AA;

  Query Match             2.4%;  Score 176;  DB 6;  Length 1555;
  Best Local Similarity   19.2%;  Pred. No. 0.00028;
  Matches  229;  Conservative  144;  Mismatches  458;  Indels  362;  Gaps   55;

Qy          9 RNSPVEEII-----NRLNNDYNLGLQCVADTT------LTPHRRKELAESDEDFGRHDKI 57
Db        332 KNSPSEEDLAARWQAELSDD--------ADETMPTEDDLAARWQAELDDDDDDLLLDDDT 383

Qy         58 YRALNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLPWSKEPPRAATA---- 113
Db        384 TNA-----QRPPEAANIDHMN--DTSMLQSPFGTPENLARPKVQPVSYTPHQPSTSDLLS 436

Qy        114 ---GQQWALQTVLLEVLNRF--MPPPN--NTPGRTFG-RTLSG-------PSGLSRPTST 158
Db        437 GIPAQNTAAQPTNASMSSYFSAQAPPNPVTTRAESFAERSKEGYKSPYDIPEDLARPRRA 496

Qy        159 NTKRKDEPANV-TFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENL-- 215
Db        497 VANSRTVVAQPGTVPKPPPRS--SSIPAPPLKASTVS---PAPLGTSSTAPTAPQKNFFE 551

Qy        216 --------NQCTKRAKGKLSDNVAAAAAPPV-----PIASALDKVPTRRHANTRDPTATG 262
Db        552 ELPLPPPRPKSRPASSGRYTPN-APVSAPSLPQSIPPPANQYSNVPGAPQSNIGPPDPPQ 610

Qy        263 HRRADQVDSFDTSQGTSYGS-----SVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVF 317
Db        611 LQQPERLDPYSNLLAPNVPSAPAVPSTASRYSPRPPGVQAGVKPPPSPRYSPAPPQSTNA 670

Qy        318 EAGHLIESPSKGRTTKSHIDNQPLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLA 377
Db        671 VAA----------APRNRYASQPASISGQGAA-----LQFQPRTSSPLAYHEKIHYEDQG 715

Qy        378 RSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLG 437
Db        716 QSEE--RPQLQSTASP------------PPLNHSHP-------------SEQPVSSENKG 748

Qy        438 ------LKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFV-TAMTGNFESKGS 490
Db        749 PSGVDVLENVPPLSTR---------------------PQSPPKNPYAPSAYTNEFANRVA 787

Qy        491 AVV-------LSAVLDYNPDNSPTAP----------------LYLVKLKPLMFEQGCRLT 527
Db        788 PVSTGPPIAGMTGVLNSSTEESPFVPPRRSQTQSPSQTLSPRLSVPSLDP--FQRPASVH 845

Qy        528 RRFGPDRFFEILIPSPTST---SPS-----VPPVVSKQPGAVEEVIQWLTMGQHSLVGRQ 579
Db        846 GSTSPTRTVNPYAPAPVPTHNRAPSQVLEFIPPTDGQQLDSLE----------------R 889

Qy        580 WRAFFAKDAGYRKPLREF---------------QLRAEDPKPIIKE-----RVHFF---- 615
Db        890 WKG---------APIFKFGFGGAVISCFPKHIPRYSAGQAAPMIKSCPGEVRISQLNDWL 940

Qy        616 -AETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQL 674
Db        941 PAAEGIVQHPGPLKGKS--------------KKKDLVAWLSSKIAAFENANIPDFDRLSP 986

Qy        675 GLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVL-------- 726
Db        987 DASK-----------LREEKT-LLWKVIRVLVENDGVLEGSVEAQKSLRNLLFPNLQDSG 1034

Qy        727 ---GLGDV---------------PSAVQGR----------FGS-AKGMW-VIDVDDTGDE 756
Db       1035 PNQSLGDVFTPSATLQPLNAPSQPDAVDSRSVDLLRDTLVLGEREKAVWAAVDKRLWGHA 1094

Qy        757 DWIETYPSQRKWE---CDFVDKHQRTLEVRSVASELKSAGLNLQLLP-VLEDRARDKVKM 812
Db       1095 MIIASRMDRSVWQQVVQEFVRR-----EVRSATSRTESLAAFYEILAGNIEESIDELVPP 1149

Qy        813 RQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRA------------------ 853
Db       1150 SARAGLQMISKVDGH-GPAKNSLDGLDSWRETVGLVLSNRSPDDQRALVALGRLLLSYNR 1208

Qy        854 TRVSH-----GRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKS 908
Db       1209 TEAAHICFILSRVAVFGGLDDPQAN----IVLLGVDHQRLSSCAALYNDDSILLTEAYEF 1264

Qy        909 KLNIRVGRSA------YIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVAR 962
Db       1265 ATSVLAGSSVSTLPHLLAFKLIHAWSLAERGR---------KSEAQQY-----CDAIAA- 1309

Qy        963 SPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWV 1015
Db       1310 ----------ALKATTKPSGYHNQHLFF----GVDELSARLRETTSDGGSSWI 1348


RESULT 13
ABJ26240
ID   ABJ26240 standard; protein; 1832 AA.
XX
AC   ABJ26240;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene protein #898.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
```

```
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013142.
XX
PR   23-APR-2001; 2001US-0285697P.
PR   27-APR-2001; 2001US-0287066P.
PR   05-JUN-2001; 2001US-0295890P.
PR   09-JUL-2001; 2001US-0303899P.
PR   31-AUG-2001; 2001US-0316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of Aspergillus
PT   fumigatus, useful for treating or preventing infections by A. fumigatus,
PT   or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS   Disclosure; Page; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This sequence represents a protein of one of the essential genes
CC   of Aspergillus fumigatus of the invention
XX
SQ   Sequence 1832 AA;

  Query Match             2.4%;  Score 176;  DB 6;  Length 1832;
  Best Local Similarity   19.2%;  Pred. No. 0.00037;
  Matches  229;  Conservative  144;  Mismatches  458;  Indels  362;  Gaps   55;

Qy          9 RNSPVEEII-----NRLNNDYNLGLQCVADTT------LTPHRRKELAESDEDFGRHDKI 57
               :||| || :         |::|          || |       |       :  || :  |:|     |
Db        332 KNSPSEEDLAARWQAELSDD--------ADETMPTEDDLAARWQAELDDDDDDLLLDDDT 383

Qy         58 YRALNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLPWSKEPPRAATA---- 113
                  |     | ::  |     |       :      |:  |  |    |  |    |  | : :|:
Db        384 TNA-----QRPPEAANIDHMN--DTSMLQSPFGTPENLARPKVQPVSYTPHQPSTSDLLS 436

Qy        114 ---GQQWALQTVLLEVLNRF--MPPPN--NTPGRTFG-RTLSG-------PSGLSRPTST 158
                  |  | |       : : |       |||   |    :|  |: |         |  |:||
Db        437 GIPAQNTAAQPTNASMSSYFSAQAPPNPVTTRAESFAERSKEGYKSPYDIPEDLARPRRA 496

Qy        159 NTKRKDEPANV-TFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENL-- 215
                    :     |       | || ||      ||:  :  :     |  |: |  |    |:    :|
Db        497 VANSRTVVAQPGTVPKPPPRS--SSIPAPPLKASTVS---PAPLGTSSTAPTAPQKNFFE 551

Qy        216 --------NQCTKRAKGKLSDNVAAAAAPPV-----PIASALDKVPTRRHANTRDPTATG 262
                          :    : |: : |  |  :||  :     | |:     ||     :|   |
Db        552 ELPLPPPRPKSRPASSGRYTPN-APVSAPSLPQSIPPPANQYSNVPGAPQSNIGPPDPPQ 610

Qy        263 HRRADQVDSFDTSQGTSYGS-----SVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVF 317
                :: :::| :       :   |        |   |               |:  : |||        |   :|
Db        611 LQQPERLDPYSNLLAPNVPSAPAVPSTASRYSPRPPGVQAGVKPPPSPRYSPAPPQSTNA 670

Qy        318 EAGHLIESPSKGRTTKSHIDNQPLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLA 377
                |                  :: :|| |  | ||          |   :        |     |
Db        671 VAA----------APRNRYASQPASISGQGAA-----LQFQPRTSSPLAYHEKIHYEDQG 715

Qy        378 RSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLG 437
                :||| |  |:|   |                           |    |                    :  :  |: |
Db        716 QSEE--RPQLQSTASP------------PPLNHSHP-------------SEQPVSSENKG 748

Qy        438 ------LKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFV-TAMTGNFESKGS 490
                      |:  |  ||                                  |: || :  : :|  |   | :: :
Db        749 PSGVDVLENVPPLSTR--------------------PQSPPKNPYAPSAYTNEFANRVA 787

Qy        491 AVV-------LSAVLDYNPDNSPTAP---------------LYLVKLKPLMFEQGCRLT 527
                |              :: ||: :  :  ||   |                       | :   |  |   |:: :
Db        788 PVSTGPPIAGMTGVLNSSTEESPFVPPRRSQTQSPSQTLSPRLSVPSLDP--FQRPASVH 845

Qy        528 RRFGPDRFFEILIPSPTST---SPS-----VPPVVSKQPGAVEEVIQWLTMGQHSLVGRQ 579
                  |  |           |:|   |    :||    :||    :|   ::|                    :
Db        846 GSTSPTRTVNPYAPAPVPTHNRAPSQVLEFIPPTDGQQLDSLE---------------R 889

Qy        580 WRAFFAKDAGYRKPLREF---------------QLRAEDPKPIIKE-----RVHFF---- 615
                |:               |: :|                    :  |       |:||      |:
Db        890 WKG---------APIFKFGFGGAVISCFPKHIPRYSAGQAAPMIKSCPGEVRISQLNDWL 940

Qy        616 -AETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQL 674
                 ||   |    |    | :|                      |         :: ||              :: |  |:
Db        941 PAAEGIVQHPGPLKGKS--------------KKKDLVAWLSSKIAAFENANIPDFDRLSP 986

Qy        675 GLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVL-------- 726
                  ||            :|  || ||               | ||||    |       | :|::|
Db        987 DASK-----------LREEKT-LLWKVIRVLVENDGVLEGSVEAQKSLRNLLFPNLQDSG 1034

Qy        727 ---GLGDV---------------PSAVQGR----------FGS-AKGMW-VIDVDDTGDE 756
                    ||||                 | ||  |                 |      | :|  :|     |
Db       1035 PNQSLGDVFTPSATLQPLNAPSQPDAVDSRSVDLLRDTLVLGEREKAVWAAVDKRLWGHA 1094

Qy        757 DWIETYPSQRKWE---CDFVDKHQRTLEVRVSASELKSAGLNLQLLP-VLEDRARDKVKM 812
                 |  :   |:     :|| :        ||||  |      :  :|     :|:   :  |
Db       1095 MIIASRMDRSVWQQVVQEFVRR-----EVRSATSRTESLAAFYEILAGNIEESIDELVPP 1149

Qy        813 RQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRA------------------ 853
                   | ::|: :        |::|:  :      :| ||  |:|:
Db       1150 SARAGLQMISKVDGH-GPAKNSLDGLDSWRETVGLVLSNRSPDDQRALVALGRLLLSYNR 1208

Qy        854 TRVSH-----GRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKS 908
                | :|      ||     |||  | |            ::  | | ::         :   |         :  :
Db       1209 TEAAHICFILSRVAVFGGLDDPQAN----IVLLGVDHQRLSSCAALYNDDSILLTEAYEF 1264

Qy        909 KLNIRVGRSA------YIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVAR 962
                 :: |  |              :  :      | : |                :  |  :  :          || : |
Db       1265 ATSVLAGSSVSTLPHLLAFKLIHAWSLAERGR---------KSEAQQY-----CDAIAA- 1309

Qy        963 SPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWV 1015
                          ::|  ||  :  :          |       |: : :|         ||  :|:
Db       1310 ----------ALKATTKPSGYHNQHLFF----GVDELSARLRETTSDGGSSWI 1348

RESULT 14
ABB65616
ID   ABB65616 standard; protein; 2271 AA.
XX
AC   ABB65616;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 23640.
XX
```

```
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL09719.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 23640; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2271 AA;

  Query Match             2.3%;  Score 173.5;  DB 4;  Length 2271;
  Best Local Similarity   18.7%;  Pred. No. 0.00087;
  Matches  191;  Conservative  147;  Mismatches  395;  Indels  287;  Gaps   44;

Qy         58 YRALNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHA------DPDTLPWSKEPPRAA 111
              : |||: ||:|   : :  |  || |   ::    |::      :|| | ||    |  :
Db        935 FELLNFIKLRKNDGDIRYQINVKIEWKKVLAHHNKMANSKTRIIIEPDCLRWS---PLLS 991

Qy        112 TAGQQWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPGSGLSRPTSTNTKRKDE------ 165
               :       |  : :  ||                  :|          : :  ||::|
Db        992 VS----KLPNIIKPISNREYLSNQMETNSSFKEINEDKISVKQGVSTSSKVNQESSELST 1047

Qy        166 ----PANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQCTKR 221
                  |::   |   :   |               | | :|   :  :|:| | :   |    :
Db       1048 KNRGPSSEKLIDERENQRKR----------AYPEEFLKSASNEAKKPKLTAPQPN-LEDQ 1096

Qy        222 AKGKLSDNVAAAAAPPVP---------IASALDKVPTRRHANTRDPTATGHRRADQVDSF 272
              : |:|||:    :: :              :|  |:|   | |  |                |
Db       1097 SFGELSDSSDCKSSKDLSEQLTKRAQRLAKRNDLIP---HTNKR-------------KHF 1140

Qy        273 DTSQGTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPS----- 327
              |              | :: | |:    ||:|  :       |    | || |  :
Db       1141 D---------------RPLENNTTSTDHNPPAQSQ----ADTAGTENGHKSEETAAKENC 1181

Qy        328 KGRTTKSHIDNQPLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQV 387
               |    | |:| :|::                |  · |:         :|||  ||||   ::
Db       1182 KNVVEKEHVD-KPVN--------------IMPKLRGK-------QSNG-KRSEEPQFNEN 1218

Qy        388 QVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLG-----LKYDP 442
                 || |  |         | :   |: ||                  : ||  |       |:|
Db       1219 PTSDPVAVASLSP-----AKEITREPVKVA--------------KTESPGKNQESAKHDA 1259

Qy        443 SWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNP 502
               :: |         |||        :|     :|     : | |           : ||
Db       1260 EPNSVSDQV--------LDA-----VAKKTKKTLFSDVASYNTED---------VKYNS 1296

Qy        503 DNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVE 562
              : |  |     |   |  | | : |:          |:        : ||  |  |::
Db       1297 EEKPMAIAQEPKESPT--EAGIKETKP------LEVPEQKAKAKSPEKPATVAE------ 1342

Qy        563 EVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITF 622
                     |: ||                  |:   ::  |: |   ::   ||:      :
Db       1343 ------TVLQH---------------VHAPVNTEKISTEESK--LELDVHY-----LKL 1373

Qy        623 RPDVFKTRSVVPAEEPV-------EQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLG 675
               ||       |     ::||     |:  :  : :|   : |  :      || :::  : :
Db       1374 SPDSALNPPVASVQKPVPEPPAPKEKAKDLQAAQPSEALQK------QPEIEMVKKTE-- 1425

Qy        676 LSKTYAIMTLEPHQIRHHKTDLLS---------PSGTGEVM-NDGVGRMSRSVAKRIRDV 725
              :||   ::  |      ::||: |          |      :   |   :   ::: |  :
Db       1426 -TKTETPVSNPP----SNRTDVSSLVEQPVKAAPEKVEQPMVKEAVDIKDKALKKPV--- 1477

Qy        726 LGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSV 785
                    |||               |:  :  |   |   |:   ::    | : :     |: ::|:
Db       1478 ---PDVP-------------VVKPEATSCEKKIDHSKTKVFHEKEAIKTDQQLVQVKE- 1519

Qy        786 ASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWV 845
                 |   :|         :|:      |||:::        :|  |  :  ||    :         ::  ||
Db       1520 -EEKSNARAPSAPIPI-----RDKIQLKGNEHAQLQNSIPSQFPLNQMPNYHTSQYWQWE 1573

Qy        846 YESYS-SRATRVSHGRVPFLAGLPDSQEETLNFLMN------------SGFDPKKQKYL 891
                |  |:  |          |:     |     :    |||  |                 |:|:
Db       1574 YYGYNLSHLDASQKGQKQFHKDLATTMAYTHNFTQNLYQSANLAMQHAHHQMHQTKEKHK 1633

Qy        892 QDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFT 951
               :         ||     | : | :|  |   :    |                : ||   :::    ::
Db       1634 VE-------RKNSGKKEEQN-KVVSSVNVVSTA-------REDAHVQHCNEYAANNQAAL 1678

Qy        952 LLSDC-----DVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGG 1006
                              :|  |: :  : |          ||          :  |   | || :|: |
Db       1679 YNQKCASQQKQAQQMKSLANVQNPVPRQSNSNAPESTVLMPNPVISQAGSVPAKQKVEHG 1738

RESULT 15
ABR52986
ID   ABR52986 standard; protein; 1226 AA.
XX
AC   ABR52986;
XX
DT   20-JUN-2003  (first entry)
XX
DE   Protein sequence #SEQ ID 837.
XX
KW   Multiprotein complex; eukaryote; drug target; diagnosis.
XX
OS   Saccharomyces cerevisiae.
XX
PN   EP1258494-A1.
XX
PD   20-NOV-2002.
XX
PF   20-DEC-2001; 2001EP-00130253.
XX
PR   15-MAY-2001; 2001EP-00111774.
XX
PA   (CELL-) CELLZOME AG.
XX
PI   Bauer A,  Gavin A,  Grandi P,  Krause R,  Kruse UD,  Kuester BD;
PI   Marzioch M,  Schultz JD,  Superti-Furga GD;
XX
DR   WPI; 2003-250078/25.
DR   N-PSDB; ACC61028.
XX
PT   New isolated protein complexes useful for diagnosing a disease or
PT   disorder, or as a target for an active agent of a pharmaceutical,
PT   preferably a drug target in the treatment or prevention of disease or
```

```
PT   disorder.
XX
PS   Disclosure; SEQ ID NO 837; 17pp + Sequence Listing; English.
XX
CC   The invention relates to multiprotein complexes from eukaryotes. Proteins
CC   of the invention and DNA sequences encoding them are given in records
CC   ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
CC   obtainable by using a protein as a bait and isolating the set of proteins
CC   which is attached thereto from cells. Such protein complexes may comprise
CC   up to 30 distinct proteins. Protein complexes of the invention are useful
CC   for diagnosing a disease or disorder, or as a target for an active agent
CC   of a pharmaceutical, preferably a drug target in the treatment or
CC   prevention of a disease or disorder. Note: The sequence data for this
CC   patent is not represented in the printed specification, but is based on
CC   sequence information supplied by the European Patent Office. The complete
CC   document is available on CD-ROM
XX
SQ   Sequence 1226 AA;

  Query Match             2.3%;  Score 168;  DB 6;  Length 1226;
  Best Local Similarity   18.8%;  Pred. No. 0.00085;
  Matches  221;  Conservative  174;  Mismatches  466;  Indels  314;  Gaps   50;

Qy         84 KAASSNWVPKAHADPDTLPWSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFG 143
              |:::: :|     ||  |  |  |   |  |:                   :|  |   || |||
Db         38 KSSNNGFVSSPTADNSTNP-SVTPSTAS-------------------VPLPTAAPGSTFG 77

Qy        144 -----RTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFP 198
                   :   || :|||:|::     :    :     |   :  |  |  |:|      :: :  |
Db         78 IEAPRPSRYDPSSVSRPSSSSYSSTRKIGSRYNPDVERSSSTTSSTPESMNTSTITHTNT 137

Qy        199 DPVNTGSKRPSLESENLNQCTKRAKGKLSDNV-----AAAAAPPVPIASALDKVPTRRHA 253
              |   |:   | ::    |           :  | ||     | : |||  :|:    :  |  |
Db        138 DIGNSRYSRKTMSRYN-------PQSTSSTNVTHFPSALSNAPPFYVANGSSRRP--RSM 188

Qy        254 NTRDPTATGHRRADQVDSFDTSQGTSYGS------SVFSACR-------HNQSTTQSSFE 300
               :   |   |       :  |  |  :  :   ||  |      |:  :  |          |  :||  ::
Db        189 DDYSPDVTNKLETNNVSSVNNNSPHSYYSRSNKWRSIGTPSRPPFDNHVGNMTTTSNTNS 248

Qy        301 APPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQ--------PLSSSSQGETSFS 352
                 :|   |   |: ::  |    |||   |  |  |   |:             :  :  |:  ||
Db        249 IHQREPFWKAN-STTILKSTHSQSSPSL-HTKKFHDANKLDKPEASVKVETPSKDETKAI 306

Qy        353 TYYE-SFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWL-- 409
              :|:: :||          |    :  :     ||  |    :  :||          |  |   :  |   :
Db        307 SYHDNNFPPRKSVSKPNAPLEPDNIKVGEEDALGKKEVHKSG-----REIAKEHPTPVKM 361

Qy        410 --HEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKP 467
                |:   |  |  :|:::              ::   |   |  :||:|:  |               |
Db        362 KEHDELEARAKKVSKI-----------NIDGKQDEIWTTAKTVA---------SAVEVSK 401

Qy        468 FPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKL----------KP 517
                :|          :|  :  |  |  |  |    :        |:|      |      :  ||             :|
Db        402 ESQK--------ELTRSVERKESPEIRDYERAYDPKALKTD---VTKLTVDNDNKSYEEP 450

Qy        518 LMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQ--HSL 575
              |    :||                   |  |  |                    |       |    |   :  :
Db        451 LEKVEGC--------------IFPLPKA-----------------ETRLWELKNQKRNKI 479

Qy        576 VGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPA 635
              :   |               :|  :|  |    :|  |        :  :|    |  ||:     :   |    |   :
Db        480 ISEQKYLL-------KKAIRNF---SEYPFYAQNKLIHQQA-TGLI----LTKIISKIKK 524

Qy        636 EEPVEQRTEFKVSQMLDWL-LQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHK 694
              || ::      |::     |:   ||        ::     :: :::      |  |             :|::    |
Db        525 EEHLK-----KINLKHDYFDLQ---KKYEKECEILTKLSENLRKEEIENKRKEHELMEQK 576

Qy        695 TDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTG 754
                           :|:             |  :|            ||:
Db        577 R-----------REEGI---ETEKEKSLRH-------PSS-------------------- 595

Qy        755 DEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRAR------- 807
                       :   |:|:    ||||   :     :     :          |           :  |::  |    |
Db        596 ------SSSSRRRNRADFVDDAEMENVLLQIDPNYKHYQAAATIPPLILDPIRKHSYKFC 649

Qy        808 ---DKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFL 864
                   : |   ::      |::  |      |::  :|:|       |   :   |    :   :   ::||       ::
Db        650 DVNNLVTDKKLWASRILKDASDNFTDHEHSL-----FLEG-YLIHPKKFGKISH----YM 699

Qy        865 AGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIA 924
                || :|   |::        :   ||     :  :   |   |::   :   :|     |   ||
Db        700 GGLRSPEECVLHY-----YRTKKTVNYKQLLIDKNKKRKMSAAAKRRKRKERS------- 747

Qy        925 DFWGVLEENEVHVGFSSK-----FRDEEESFTLLSD--CDVLVARS---------PAHFP 968
                     : ||  |   |  :              ||:|          :          |||      |           |
Db        748 ------NDEEVEVDESKEESTNTIEKEEKSENNAEENVQPVLVQGSEVKGDPLGTPEKVE 801

Qy        969 SDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPE--IVDGFV 1026
              : |::       |   || : :  |             |          :               : :         ||:   : :|:
Db        802 NMIEQRGEEFAGELENAERVNDLKRAHDEVGEESNKSSVIETNNGVQIMDPKGAVQNGYY 861

Qy       1027 NAE-----MPLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHF------ 1075
                  |          ||  |  |  ||  |                          || : :|:|   :::|   |
Db        862 PEETKELDFSLENALQR--KKHK-------------SAPEHKTSYWSVRESQLFPELLKE 906

Qy       1076 -----ALQPNFLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWA 1130
                    :|     ||   :        |     |    |:     |       :::                  |    :     |                  :
Db        907 FGSQWSLISEKLGTKSTTMVRNYYQRNAARNGWKLLVDE-----TDLKRDGTSSESVQQS 961

Qy       1131 QL----RRELLGGALSLPDPMYKSDSWLGRGEPTH 1161
              |:       |    :        ::|  |     :           |:|||
Db        962 QILIQPERPNINAYSNIP-PQQRPALGYFVGQPTH 995

Search completed: March 29, 2004, 08:47:38
Job time : 73 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 29, 2004, 08:46:22 ; Search time 26 Seconds (without alignments) 2783.831 Million cell updates/sec

Title: US-09-913-878A-2
Perfect score: 7397
Sequence: 1 MNPITPRKRNSPVEEIINRL..........YEVLGDDDFDGIGFTGNGDY 1402

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	297	4.0	1114	3	US-08-811-583-2	Sequence 2, Appli
2	151.5	2.0	218	3	US-08-811-583-3	Sequence 3, Appli
3	148.5	2.0	1457	4	US-09-436-874-2	Sequence 2, Appli
4	142.5	1.9	1457	4	US-09-713-273A-18	Sequence 18, Appl
5	142	1.9	2409	6	5180808-2	Patent No. 5180808
6	141.5	1.9	909	3	US-09-425-383-2	Sequence 2, Appli
7	139.5	1.9	1312	3	US-09-041-886-19	Sequence 19, Appl
8	139.5	1.9	1312	4	US-09-648-281-2	Sequence 2, Appli
9	139.5	1.9	1312	4	US-09-707-919A-19	Sequence 19, Appl
10	139.5	1.9	1312	4	US-09-083-268-3	Sequence 3, Appli
11	139.5	1.9	1384	3	US-08-976-255-11	Sequence 11, Appl
12	139	1.9	1528	4	US-09-845-917A-3	Sequence 3, Appli
13	139	1.9	1583	4	US-09-845-917A-4	Sequence 4, Appli
14	137	1.9	1768	4	US-09-489-039A-11704	Sequence 11704, A
15	136.5	1.8	2972	4	US-09-579-181-2	Sequence 2, Appli
16	136.5	1.8	3118	4	US-09-579-181-1	Sequence 1, Appli
17	129.5	1.8	1317	3	US-09-083-521-7	Sequence 7, Appli
18	128	1.7	1093	4	US-09-252-991A-21827	Sequence 21827, A
19	128	1.7	3969	3	US-08-061-376-5	Sequence 5, Appli
20	127.5	1.7	1191	4	US-09-921-099A-2	Sequence 2, Appli
21	127.5	1.7	1191	4	US-09-921-099A-4	Sequence 4, Appli
22	127	1.7	1341	4	US-09-252-991A-26785	Sequence 26785, A
23	125.5	1.7	511	4	US-09-252-991A-26078	Sequence 26078, A
24	125.5	1.7	3418	2	US-08-639-501-2	Sequence 2, Appli
25	125.5	1.7	3418	3	US-09-044-946-2	Sequence 2, Appli
26	125.5	1.7	3418	3	US-08-755-587-44	Sequence 44, Appl
27	125.5	1.7	3418	3	US-09-044-908-2	Sequence 2, Appli
28	124.5	1.7	2188	4	US-09-328-352-7763	Sequence 7763, Ap
29	124.5	1.7	3418	2	US-08-603-753D-4	Sequence 4, Appli
30	124.5	1.7	3418	3	US-09-099-753-4	Sequence 4, Appli
31	124.5	1.7	3418	3	US-08-986-106-4	Sequence 4, Appli
32	124	1.7	1581	3	US-09-110-517-2	Sequence 2, Appli
33	122.5	1.7	801	4	US-09-252-991A-27870	Sequence 27870, A
34	122.5	1.7	1742	4	US-09-386-962C-4	Sequence 4, Appli
35	122	1.6	3898	3	US-08-750-717-2	Sequence 2, Appli
36	121.5	1.6	582	4	US-09-450-072-79	Sequence 79, Appl
37	121.5	1.6	582	4	US-09-351-348-79	Sequence 79, Appl
38	121	1.6	521	2	US-08-682-847-4	Sequence 4, Appli
39	121	1.6	1394	4	US-09-845-917A-57	Sequence 57, Appl
40	121	1.6	1449	4	US-09-845-917A-58	Sequence 58, Appl
41	120	1.6	889	4	US-09-252-991A-18702	Sequence 18702, A
42	119.5	1.6	525	4	US-10-132-920B-27	Sequence 27, Appl
43	119.5	1.6	1130	4	US-09-442-100-4	Sequence 4, Appli
44	119.5	1.6	1130	4	US-09-233-857-3	Sequence 3, Appli
45	119.5	1.6	1130	4	US-08-939-106-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-08-811-583-2
; Sequence 2, Application US/08811583
; Patent No. 6218142
;  GENERAL INFORMATION:
;    APPLICANT:  Wassenegger, Michael
;    APPLICANT:  Riedel, Leonhard
;    APPLICANT:  Schiebel, Winfried
;    APPLICANT:  Sanger, Heinz
;    TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING
;    TITLE OF INVENTION:  POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
;    TITLE OF INVENTION:  RNA-DIRECTED RNA POLYMERASE (RdRP)
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & NEAVE
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/811,583
;      FILING DATE:  05-MAR-1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley, James F.
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  MPG-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-596-9000
;      TELEFAX:  212-596-9090
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1114 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-811-583-2

  Query Match            4.0%;  Score 297;  DB 3;  Length 1114;
  Best Local Similarity  20.4%;  Pred. No. 2.5e-16;
  Matches  175;  Conservative  133;  Mismatches  323;  Indels  228;  Gaps   32;

Qy        562 EEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGIT 621
              |::  :| :   | |  :|   :: |                  ||           | ||
```

```
Db        386 EDIDNFLRV---SFVDEEWEKLYSTDL--------------LPK----------ASTGSG 418

Qy        622 FRPDVFK------TRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTW----QPHLKL--- 668

Db        419 VRTNIYERILSTLRKGFVIGDKKFE-FLAFSSSQL------RDNSVWMFASRPGLTANDI 471

Qy        669 ------FSRI--------QLGLSKTYAIMTLEPHQIRHHKTDLL---SPSGTGEVMNDGV 711

Db        472 RAWMGDFSQIKNVAKYAARLGQSFGSSRETL---SVLRHEIEVIPDVKVHGTSYVFSDGI 528

Qy        712 GRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECD 771

Db        529 GKISGDFAHRVASKCGLQYTPSAFQIRYGGYKGVVGVDPDSSMK---LSLRKSMSKYESD 585

Qy        772 FVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQ 831

Db        586 NI-----KLDVLG-WSKYQPCYLNRQLITLLS---------TLGVKDEVLEQKQKEAVDQ 630

Qy        832 KHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYL 891

Db        631 LDAILHDSLKAQEALELMS-----------------PGENTNILKAMLNCGYKPDAEPFL 673

Qy        892 QDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFS----SKFRDEE 947

Db        674 SMMLQTFRASKLLDLRTRSRIFIPNGRTMMGCLDESRTLEYGQVFVQFTGAGHGEFSDDL 733

Qy        948 ESF--------TLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPL 999

Db        734 HPFNNSRSTNSNFILKGNVVVAKNPCLHPGDIRVLKAVNVRALHHMVDCVVFPQKGKRPH 793

Qy       1000 AKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSA 1059

Db        794 PNECSGSDLDGDIYFVCWDQDMIPPRQVQPMEYPPAPSIQLDHD---------------- 837

Qy       1060 AKEQTTYDMIQKSFHFALQPNFLGMCTN----YKERLCYINNSVSNKPAIILSSLVGNLV 1115

Db        838 ----VTIEEVEEYFTNYIVNDSLGIIANAHVVFADR---EPDMAMSDPCKKLAELFSIAV 890

Qy       1116 DQSKQGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAI 1175

Db        891 DFPKTGVPAEIPS--QLRPK------EYPDFMDKPD------KTSYISERVIGKLFRKVK 936

Qy       1176 DKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEV- 1234

Db        937 DKA---------------------PQASSIATFTRDVARRSYDADMEVDGFEDYIDEAF 974

Qy       1235 --EKEYGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDKSGANYD--------SKV 1284

Db        975 DYKTEY-----DNKLGNLMDYYGIKTE------AEILSGGIMKASKTFDRRKDAEAISVA 1023

Qy       1285 IRLLELSFLADREMNTW---------ALLRASTAFKLYYHKS--------------PKFV 1321

Db       1024 VRALR------KEARAWFKRRNDIDDMLPKASAWYHVTYHPTYWGCYNQGLKRAHFISFP 1077

Qy       1322 WQMAGRQLAYIKAQMTSRP 1340

Db       1078 WCVYDQLIQIKKDKARNRP 1096
```

RESULT 2
US-08-811-583-3
; Sequence 3, Application US/08811583
; Patent No. 6218142
; GENERAL INFORMATION:
; APPLICANT: Wassenegger, Michael
; APPLICANT: Riedel, Leonhard
; APPLICANT: Schiebel, Winfried
; APPLICANT: Sanger, Heinz
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RdRP)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,583
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MPG-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-583-3

```
  Query Match             2.0%;  Score 151.5;  DB 3;  Length 218;
  Best Local Similarity   24.9%;  Pred. No. 7.2e-05;
  Matches   52;  Conservative   34;  Mismatches   84;  Indels   39;  Gaps    5;

Qy        930 LEENEVHVGFS----SKFRDEEESF--------TLLSDCDVLVARSPAHFPSDIQRVRAV 977

Db         13 LEYGQVFVQFTGAGHGEFSDDLHPFNNSRSTNSNFILKGNVVVAKNPCLHPGDIRVLKAV 72

Qy        978 FKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLS 1037

Db         73 NVRALHHMVDCVVFPQKGKRPHPNECSGSDLDGDIYFVCWDQDMIPPRQVQPMEYPPAPS 132

Qy       1038 RYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTN----YKERLC 1093

Db        133 IQLDHD-------------------VTIEEVEEYFTNYIVNDSLGIIANAHVVFADR-- 170

Qy       1094 YINNSVSNKPAIILSSLVGNLVDQSKQGI 1122

Db        171 -EPDMAMSDPCKKLAELFSIAVDFPKTGV 198
```

RESULT 3
US-09-436-874-2
; Sequence 2, Application US/09436874
; Patent No. 6521816
; GENERAL INFORMATION:
; APPLICANT: Frohberg, Claus
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES FROM RICE AND THEIR USE FOR THE
; TITLE OF INVENTION: PRODUCTION OF MODIFIED STARCH
; FILE REFERENCE: GFB-10
; CURRENT APPLICATION NUMBER: US/09/436,874
; CURRENT FILING DATE: 1999-11-09
; EARLIER APPLICATION NUMBER: US 60/107,883
; EARLIER FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1457
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-436-874-2

```
Query Match             2.0%;  Score 148.5;  DB 4;  Length 1457;
Best Local Similarity   18.0%;  Pred. No. 0.004;
Matches  267;  Conservative  213;  Mismatches  532;  Indels  471;  Gaps   69;

Qy         95 HADPDTLPWSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSR 154
Db         21 HARPASPSPALLPPAALRRGRRLPAATTTLAVSRRSLLAP---------RAIAASTGRAS 71

Qy        155 P--TSTNTKRKDEPANVTFADPPKRSL-------TRSATGPPIHGAAIPLKFPD------ 199
Db         72 PGLVGRFTLDANSELKVTLNPAPQGSVAEINLEATNTSGSLILHWGALR---PDRGEWLL 128

Qy        200 PVNTGSKRPSLESENLNQCTKRAKGKLSDNVAAAAAPPVPIASAL-----DKVPTRRHAN 254
Db        129 P----SRKPDGTTVYKNRALRTPFIKSGDNSTLKIEIDDPAVQAIEFLIFDEARNNWYKN 184

Qy        255 TRDPTATGHRRADQVDSFD-TSQGTSYGSS----------VFSACRHNQSTTQSSFEAPP 303
Db        185 ------NGQNFQIQLQASQYQGQGTSTATSSTVVPEDLVQIQSYLRWERKGKQS---YTP 235

Qy        304 SQPREKRPVDATVFEAG--HLIESPSKG--------RTTKSHIDNQPLSSSSQGETSFST 353
Db        236 EQEKEE-------YEAARTELIEELNKGVSLEKLRAKLTKT------------------ 269

Qy        354 YYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKF--PKWLHE 411
Db        270 --------------PEATDSNAPA-SESTVTTKVPEELVQVQAYIR--WEKAGKPNYAPE 312

Qy        412 APLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEK 471
Db        313 KQL-------------VEFEEARKELQSELDKGTS------------VEQLRNK----- 341

Qy        472 PPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFG 531
Db        342 --------ILKGNIETKVSK---------------------QLKDKKYFSVERIQRK-- 369

Qy        532 PDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYR 591
Db        370 -KRDIVQLLKKHKPTVMEAQAETPKQPTVLD-------------------LFTK----- 403

Qy        592 KPLREFQLRAEDPKPIIKERVHFFAE---TGITFRPDVFKTRSVVPAEEPVEQRTEFKVS 648
Db        404 ------SLQEQDNCEVLSRKLFKFGDKEILGIT---------TVALGKTKVHLATNYMEP 448

Qy        649 QMLDWLLQLDNNTWQ--PHLKL--------------FSRIQL-GLSKTYAIMTLEPHQIR 691
Db        449 LILHWALSKENGEWQAPPSSILPSGSSLLDKACETSFSEYELNGLH--CQVVEIELDDGG 506

Qy        692 HHKTDLLSPSGTGEVMNDGVG---RMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVI 748
Db        507 YKRMPFVLRSGETWMKNNGSDFYLDFSTKVAKNTKDT---GDA-----GK-GTAKAL--- 554

Qy        749 DVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVR-SVASELKSAGLNLQLLPVLEDRAR 807
Db        555 ----------LERI-------ADLEEDAQRSLMHRFNIAAD-------------LVDQAR 584

Qy        808 DK-------------VKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRA 853
Db        585 DNGLLGIIGIFVWIRFMATRQLIWNKNYNVKPREISKAQDRFTDDLENMYRTYPQY---- 640

Qy        854 TRVSHGRVPFLAGLPDSQEETLNFLMNS---GFDPKKQKYLQDIAWDLQK---------- 900
Db        641 -----------------QEILRMIMSAVGRGGEGDVGQRIRDEILVIQRNNDCKGGMME 682

Qy        901 ---RKCDTLKSKLNIRVGRSAYIYMIADF-----WGVLEENEV----------HVGFSSK 942
Db        683 EWHQKLHNNTSPDDVVICQALLDYIKSDFDIGVYWDTLKKDGITKERLLSYDRPIHSEPN 742

Qy        943 FRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAV---FKPELHSLKDVIIFSTKGDVPL 999
Db        743 FRSEQKDGLLRDLGNYMRSLKAVHSGADLESAIATCMGYKSEGEG------FMVGVQINP 796

Qy       1000 AKKLSGG---------DYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQL 1050

Db        797 VKGLPSGFPKLLEFVLDHVEDKS----AEPLLEGLLEARAELHPLLLGSPERMKDLIFLD 852

Qy       1051 MASHGTGSAAKEQTTYDMI-----QKSFHF-ALQPNFLGMCTNYKERLCY---------- 1094
Db        853 IALDSTFRTAVER-SYEELNNVEPEKIMYFISLVLENLALSTDDNEDILYCLKGWNQALE 911

Qy       1095 INNSVSNKPAIILSSLVG--------------NLVDQSKQ--GIVFNEASWAQ--LRREL 1136
Db        912 MAKQKNNQWALYAKAFLDRTRLALASKGEQYYNLMQPSAEYLGSLLNIDQWAVNIFTEEI 971

Qy       1137 LGGA--------LSLPDPMYKSDSWLGRGE---PTHIIDYL-----KFSIARPAIDKELE 1180
Db        972 IRGGSAATLSALLNRIDPVLRNVAQLGSWQVISPVEVSGYIVVVDELLAVQNKSYDKPTI 1031

Qy       1181 AFHNAMKAAKDTEDG-AHFWDPDLASYYTFFKEISDKSRSSALLFTTL--KNRIGEVEKE 1237
Db       1032 LVAKSVKGEEEIPDGVVGVITPDMPD---VLSHVSVRARNCKVLFATCFDPNTLSELQGH 1088

Qy       1238 YGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDKSGANYDSKVIRLLELSFLADRE 1297
Db       1089 DGKVFSFK---------PTSADITYRE----IPESELQSGS-----------LNAEAGQA 1124

Qy       1298 MNTWALLRASTAFKLYYHKSPKFVWQMAG---RQLAYIKAQMTSRPGEGAPALMTAFMYA 1354
Db       1125 VPSVSLVKKKFLGK-YAISAEEFSEEMVGAKSRNVAYLKGKVPSWVGVPTSVAIPFGTFE 1183

Qy       1355 GLMPDK--KFTKQYVARLEGDGSEYPDPEVYEVLGDDDFDGIG 1395
Db       1184 KVLSDEINKEVAQTIQMLKGK-------------LAQDDFSALG 1214
```

RESULT 4
US-09-713-273A-18
; Sequence 18, Application US/09713273A
; Patent No. 6620987
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Butler, Karla
; TITLE OF INVENTION: STARCH R1 PHOSPHORYLATION PROTEINS
; FILE REFERENCE: BB1158 US CIP
; CURRENT APPLICATION NUMBER: US/09/713,273A
; CURRENT FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: 60/081,143
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: PCT/US99/07639
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 09/679,933
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
;   LENGTH: 1457
;   TYPE: PRT
;   ORGANISM: Oryza sativa
US-09-713-273A-18

```
  Query Match             1.9%;  Score 142.5;  DB 4;  Length 1457;
  Best Local Similarity   18.1%;  Pred. No. 0.013;
  Matches  268;  Conservative  212;  Mismatches  532;  Indels  471;  Gaps   70;

Qy         95 HADPDTLPWSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSR 154
Db         21 HARPASPSPALLPPAALRRGRRLPAATTTLAVSRRSLLAP---------RAIAASTGRAS 71

Qy        155 P--TSTNTKRKDEPANVTFADPPKRSL-------TRSATGPPIHGAAIPLKFPD------ 199
Db         72 PGLVGRFTLDANSELKVTLNPAPQGSVVEINLEATNTSGSLILHWGALR---PDRGEWLL 128

Qy        200 PVNTGSKRPSLESENLNQCTKRAKGKLSDNVAAAAAPPVPIASAL-----DKVPTRRHAN 254
Db        129 P----SRKPDGTTVYKNRALRTPFIKSGDNSTLKIEIDDPAVQAIEFLIFDEARNNWYKN 184
```

```
Qy    255 TRDPTATGHRRADQVDSFD-TSQGTSYGSS----------VFSACRHNQSTTQSSFEAPP 303
Db    185 ------NGQNFQIQLQASQYQGQGTSTATSSTVVPEDLVQIQSYLRWERKGKQS---YTP 235

Qy    304 SQPREKRPVDATVFEAG--HLIESPSKG--------RTTKSHIDNQPLSSSSQGETSFST 353
Db    236 EQEKEE-------YEAARTELIEELNKGVSLEKLRAKLTKT------------------- 269

Qy    354 YYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKF--PKWLHE 411
Db    270 -------------PEATDSNAPA-SESTVTTKVPEELVQVQAYIR--WEKAGKPNYAPE 312

Qy    412 APLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEK 471
Db    313 KQL-------------VEFEEARKELQSELDKGTS------------VEQLRNK----- 341

Qy    472 PPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFG 531
Db    342 --------ILKGNIETKVSK---------------------QLKDKKYFSVERIQRK-- 369

Qy    532 PDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYR 591
Db    370 -KRDIVQLLKKHKPTVMEAQVETPKQPTVLD--------------------LFTK----- 403

Qy    592 KPLREFQLRAEDPKPIIKERVHFFAE---TGITFRPDVFKTRSVVPAEEPVEQRTEFKVS 648
Db    404 ------SLQEQDNCEVLSRKLFKFGDKEILGIT---------TVALGKTKVHLATNYMEP 448

Qy    649 QMLDWLLQLDNNTWQ--PHLKL---------------FSRIQL-GLSKTYAIMTLEPHQIR 691
Db    449 LILHWALSKENGEWQAPPSSILPSGSSLLDKACETSFSEYELNGLH--CQVVEIELDDGG 506

Qy    692 HHKTDLLSPSGTGEVMNDGVG---RMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVI 748
Db    507 YKRMPFVLRSGETWMKNNGSDFYLDFSTKVAKNTKDT---GDA-----GK-GTAEAL--- 554

Qy    749 DVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVR-SVASELKSAGLNLQLLPVLEDRAR 807
Db    555 ----------LERI-------ADLEEDAQRSLMHRFNIAAD-------------LVDQAR 584

Qy    808 DK-------------VKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRA 853
Db    585 DNGLLGIIGIFVWIGFMATRQLIWNKNYNVKPREISKAQDRFTDDLENMYRTYPQY---- 640

Qy    854 TRVSHGRVPFLAGLPDSQEETLNFLMNS---GFDPKKQKYLQDIAWDLQK---------- 900
Db    641 -----------------QEILRMIMSAVGRGGEGDVGQRIRDEILVIQRNNDCKGGMME 682

Qy    901 ---RKCDTLKSKLNIRVGRSAYIYMIADF-----WGVLEENEV----------HVGFSSK 942
Db    683 EWHQKLHNNTSPDDVVICQALLDYIKSDFDTGVYWDTLKKGGITKERLLSYDRPIHSEPN 742

Qy    943 FRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAV---FKPELHSLKDVIIFSTKGDVPL 999
Db    743 FRSEQKDSLLRDLGNYMRSLKAVHSGADLESAIATCMGYKSEGEG------FMVGVQINP 796

Qy   1000 AKKLSGG---------DYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQL 1050
Db    797 VKGLPSGFPKLLEFILDHVEDKS----ARPLLGGLLEARAELHPLLLGSPERMKDLIFLD 852

Qy   1051 MASHGTGSAAKEQTTYDMI-----QKSFHF-ALQPNFLGMCTNYKERLCYI----NNSV- 1099
Db    853 IALDSTFRTAVER-SYEELNNVEPEKIMYFISLVLENLALSTDDNEDILYCLKGWNQAVE 911

Qy   1100 -----SNKPAIILSSLVG--------------NLVDQSKQ--GIVFNEASWAQ--LRREL 1136
Db    912 MAKQKNNQWALYAKAFLDRTRLALASKGEQYYNLMQPSAEYLGSLLNIDQWAVNIFTEEI 971

Qy   1137 LGGA--------LSLPDPMYKSDSWLGRGE---PTHIIDYL-----KFSIARPAIDKELE 1180
Db    972 IRGGSAATLSALLNRIDPVLRNVAQLGSWQVISPVEVSGYIVVVDELLAVQNKSYDKPTI 1031

Qy   1181 AFHNAMKAAKDTEDG-AHFWDPDLASYYTFFKEISDKSRSSALLFTTL--KNRIGEVEKE 1237
Db   1032 LVAKSVKGEEEIPDGVVGVITPDMPD---VLSHVSVRARNCKVLFATCFDPNTLSELQGH 1088

Qy   1238 YGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDKSGANYDSKVIRLLELSFLADRE 1297
Db   1089 DGKVFSFK---------PTSADITYRE----IPESELQSGS-----------LNAEAGQA 1124

Qy   1298 MNTWALLRASTAFKLYYHKSPKFVWQMAG---RQLAYIKAQMTSRPGEGAPALMTAFMYA 1354
Db   1125 VPSVSLVKKKFLGK-YAISAEEFSEEMVGAKSRNVAYLKGKVPSWVGVPTSVAIPFGTFE 1183

Qy   1355 GLMPDK--KFTKQYVARLEGDGSEYPDPEVYEVLGDDDFDGIG 1395
Db   1184 KVLSDEINKEVAQTIQMLKGK------------LAQDDFSALG 1214
```

RESULT 5
5180808-2
;Patent No. 5180808
; APPLICANT: RUOSLAHTI, ERKKI I.
; TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID
;SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN
;ANTIBODIES, AND METHODS OF DETECTING THE SAME
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/441,179
; FILING DATE: 27-NOV-1989
;SEQ ID NO:2:
; LENGTH: 2409
5180808-2

```
  Query Match             1.9%;  Score 142;  DB 6;  Length 2409;
  Best Local Similarity   18.0%;  Pred. No. 0.037;
  Matches 254;  Conservative 162;  Mismatches 509;  Indels 490;  Gaps   66;

Qy    109 RAATAGQQWALQTVL-LEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNTKRKDEPA 167
Db    312 RAQCGGGLLGVRTLYRFENQTGFPPPDSRFDAYCFKRRMSDLSVIGHPIDSESK-EDEPC 370

Qy    168 NVTFADPPKRSLTRSATGPPIHG--AAIPLKFPDPVNTGSKRPSLESENLNQCTKRAKGK 225
Db    371 --------------SEETDPVHDLMAEILPEFPDIIEIDLYHSEENEEEEECAN----- 411

Qy    226 LSDNVAAAAAPPVPIASALDKVPTRRHAN--------TRDPTATGHRRADQVDSFDTSQG 277
Db    412 ---------------ATDVTTTPSVQYINGKHLVTTVPKDPEAAEARRG-QFESVAPSQN 455

Qy    278 TSYGS------------SVFSACRHNQST-TQSSFEA---PPSQPREKR-------PVDA 314
Db    456 FSDSSESDTHPFVIAKTELSTAVQPNESTETTESLEVTWKPETYPETSEHFSGGEPDVFP 515

Qy    315 TV-----FEAGHLIESPSKGRTTKSHIDNQPLSSSSQGETSFSTYYESFPSSGGEGAIPE 369
Db    516 TVPFHEEFESG----TAKKGAESVTERDTEVGHQAHEHTEPVSLFPE---ESSGEIAIDQ 568

Qy    370 PSRSNGLARS------EESARSQVQVHAPVV--------------AARLRNIWP-----K 404
Db    569 ESQKIAFARATEVTFGEEVEKSTSVTYTPTIVPSSASAYVSEEEAVTLIGNPWPDDLLST 628

Qy    405 FPKWLHEAPLAVAWEVTRLFMHCKVDL----------EDESLGLKYDPSWSTARDVTDIW 454
Db    629 KESWVEATPR----QVVELSGSSSIPITEGSGEAEEDEDTMFTMVTDLS---QRNTTD-- 679

Qy    455 KTLYRLDAFR--GKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYL 512
Db    680 -TLITLDTSRIITESFFEVPATTIYPVS------EQPSAKVVPTKFVSETDTSEWISSTT 732

Qy    513 VKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPS-VPPVVSKQPGAVEEVIQWLTMG 571
Db    733 VEEKKRKEEEGTT-----GTASTFEVY--SSTQRSDQLILPFELESPNVAT--------- 776

Qy    572 QHSLVGRQWRAFFAKDAGYRKPL------REFQLRAEDPKPIIKER---VHFFAET---- 618
```

```
Db        777 -------------SSDSGTRKSFMSLTTPTQSEREMTDSTPVFTETNTLENLGAQTTEHS 823

Qy        619 ---------GITFRP----DVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDN---NTW 662

Db        824 SIHQPGVQEGLTTLPRSPASVFMEQGSGEAAADPETTTVSSFSLNVEYAIQAEKEVAGTL 883

Qy        663 QPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRI 722

Db        884 SPHVE-------------TTFSTEP-------TGLVLSTVMDRVVAENITQTSREIV--I 921

Qy        723 RDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEV 782

Db        922 SERLGEPNYGAEIRG-------------FSTG-------FPLEEDFSGDF-------REY 954

Qy        783 RSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQ------------RQFSE 830

Db        955 STVSHPI----------------AKEETVMMEGSGDAAFRDTQTSPSTVPTSVHISHISD 998

Qy        831 QKHALNRPVEFRQWVYESYSSRA-------TRVSHGRVPFLAGLPDSQEETLNFLMNSGF 883

Db        999 SEGPSSTMVSTSAFPWEEFTSSAEGSGEQLVTVSSSVVPVLPSAVQKFSGTASSIIDEGL 1058

Qy        884 -------DPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMI---ADFWGVLEEN 933

Db       1059 GEVGTVNEIDRRSTILPTA-EVEGTKAPVEKEEVKVSGTVSTNFPQTIEPAKLWSRQEVN 1117

Qy        934 EVHVGFSSKFRDEEE-----SFTLLSDCDVLVARSPAHFPSDIQRV---RAVFKPELHSL 985

Db       1118 PVRQEIESETTSEEQIQEEKSF-----------ESPQNSPATEQTIFDSQTFTETELKT- 1165

Qy        986 KDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPD---LSRYLKK 1042

Db       1166 TDYSVLTTKKTYSDDKEMKEED---------------TSLVNMSTP-DPDANGLESYTTL 1209

Qy       1043 DKTTFK------------QLMASHGTGSA--AKEQTTYDMIQKSFHFALQPNFLGMCTNY 1088

Db       1210 PEATEKSHFFLATALVTESIPAEHVVTDSPIKKEEST-----KHFPKGMRPT-------- 1256

Qy       1089 KERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDPMY 1148

Db       1257 -------------------------IQESDTELLF----------------------- 1266

Qy       1149 KSDSWLGRGE------PTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWDPD 1202

Db       1267 ---SGLGSGEEVLPTLPTESVNF----------TEVEQINNTL---------------- 1296

Qy       1203 LASYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKD---------- 1252

Db       1297 ----YPHTSQVESTSSDKIEDF----NRMENVAKEVGPLVSQTDIFEGSGSVTSTTLIEI 1348

Qy       1253 ------------PYPVRV------NQVYEKWC----AITPEAMDKSGANYDSKVIRLLEL 1290

Db       1349 LSDTGAEGPTVAPLPFSTDIGHPQNQTV-RWAEEIQTSRPQTITEQDSNKNSSTAEINET 1407

Qy       1291 S-----FLADR---EMNTWALLRASTAFKLYYHKS 1317

Db       1408 TTSSTDFLARAYGFEMAKEFVTSAPKPSDLYYEPS 1442
```

RESULT 6
US-09-425-383-2
; Sequence 2, Application US/09425383
; Patent No. 6194637
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Shi, Jinrui
; TITLE OF INVENTION: Maize DNA Ligase I Orthologue and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0962
; CURRENT APPLICATION NUMBER: US/09/425,383
; CURRENT FILING DATE: 1999-10-22
; EARLIER APPLICATION NUMBER: 60/108,793
; EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 909
; TYPE: PRT
; ORGANISM: Zea mays
US-09-425-383-2

```
  Query Match             1.9%;  Score 141.5;  DB 3;  Length 909;
  Best Local Similarity   19.6%;  Pred. No. 0.0071;
  Matches 197;  Conservative 138;  Mismatches 361;  Indels 311;  Gaps 48;

Qy        152 LSRPTSTNT--------KRKDEPANVTFADPP---------KRSLTRSATGPPIHGAAIP 194

Db          4 LTATTSTTTATSLLLRCRPLPSPKTVTLATPPAILPFCIRSRFSASASASASSVSPQSRA 63

Qy        195 LKFPDPVNTGSKRPSLESENLNQCTKRAKGKLSDNVAAAAAPPVPIASALDKVPTRRHAN 254

Db         64 TAAAAAGKSREKKP-----------PTARAKRADAVPAAAAPGMS-ASGGGSAGKRTVAD 111

Qy        255 T--RDPTATGHRRADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKR-- 310

Db        112 VLMGNARAAASKAKKAAPSPKKPQAKTDGAEV-------EPEAAAVAEKPPSPARSKRAS 164

Qy        311 -PVDA-TVFEAGHLIESPSKGRTTKSHID----NQPLSSSSQGETSFSTYYESFPSSGGE 364

Db        165 SPAKSPKLLPDGSTSSAKSKGPAALAQLDAAKENPP---SPKRSKTLAAKSDTQPS--GE 219

Qy        365 GAI-----------PEPSRSNGLA-RSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEA 412

Db        220 GVVGLGHTDGKKGSPSPTKAKGQASQSEEKKRAASPKKAKTTDS------PKTEK----- 268

Qy        413 PLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKP 472

Db        269 ---------------NITLELKKKGSEFDPM------AVAYWK-------------PGEP 294

Qy        473 PNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGP 532

Db        295 VPFLFLARALDLISNESGRIVITEIL----------------------SNVFRTVMATTP 332

Qy        533 DRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRK 592

Db        333 D---DLLATVYLSANRIAPP---------HEGIE-LGIGDASVIRALAEAYGRKEEHVKK 379

Qy        593 PLREF-------------QLRAEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPV 639

Db        380 DLKELGDLGLVAKASRSSQKMMFKPKPLTIARVL------STFR---------TIAKESG 424

Qy        640 EQRTEFKVSQMLDWLLQLDNNTWQPHLKLF-SRIQLGLSKTYAIMTLEPHQIRHHKTDLL 698

Db        425 KDSQDKKRNHMKGLLVAATDCEPQYITRLLQSKMRIGLAEKTVQMALGQAAVYADKNPL- 483

Qy        699 SPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGD--VPSAVQGRFGSAKGMWVI-------- 748

Db        484 -PPKVQSPFEE-----AAKIIKQAYSVLPIYDKIVPAILE------VGVWKLPETCKFSI 531

Qy        749 -------------DVDDTGDEDWIETYPSQRKWECDFVDKH---QRTLEVRSVASE---- 788

Db        532 GVPVGPMLAKATKSVSEIIDKFQGLEYTCEYKYDGERAQIHCMEDGSVEIYSRNAERNTG 591

Qy        789 ----------------LKSAGLNLQLLPVLEDRARDKVKMRQAIGDR-----LINDLQRQ 827

Db        592 KYPDVVDAVSRFRKATVKSFVLDCEI--VAYDREKQKILPFQILSTRARKGVTINDIKVS 649

Qy        828 ---------FSEQKHALNRPVEF-RQWVYESYSSRATRVSHGRVPFLAGL-----PDSQE 872

Db        650 VCTFGFDILYINGKPLLQEQLKIRREHLYNSFE---------EVPGVFKLATAITSNDLE 700

Qy        873 ETLNFL---MNSGFDPKKQKYL-QDIAWDLQKRKCDTLKSKLNI--RVGRSAYIYMIADF 926

Db        701 EIQKFLDTAVNSSCEGLIIKTLDKDATYEPAKRSNNWLKLKKDYMDSIGDSLDLVPIAAF 760

Qy        927 WG--------------VLEENE-------VHVGFSSKFRDEEESFTLLSDCDVLVARSP 964
```

```
Db        761 YGRGKRTGVFGSFLLACYDEQNEEYQTICNIGTGFSEQ-QLEERSASLRSK---VIEKPK 816

Qy        965 AHF------------PSDIQRVRAV---FKPELHSLKDVIIFSTKG 995

Db        817 AYYRFGDTMNPDVWFEPSEVWEVKAADLSISP-VHRAANGIVDPNKG 862
```

RESULT 7
US-09-041-886-19
; Sequence 19, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
;   APPLICANT: Bredesen, Dale E.
;   APPLICANT: Rabizadeh, Sharroz
;   TITLE OF INVENTION: Proapoptotic Peptides, Dependence
;   TITLE OF INVENTION: Polypeptides and Methods of Use
;   NUMBER OF SEQUENCES: 72
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Campbell & Flores LLP
;   STREET: 4370 La Jolla Village Drive, Suite 700
;   CITY: San Diego
;   STATE: California
;   COUNTRY: United States
;   ZIP: 92122
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/041,886
;   FILING DATE:
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Campbell, Cathryn A.
;   REGISTRATION NUMBER: 31,815
;   REFERENCE/DOCKET NUMBER: P-LJ 2626
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (619) 535-9001
;   TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 19:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1312 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-041-886-19

```
  Query Match              1.9%;  Score 139.5;  DB 3;  Length 1312;
  Best Local Similarity   17.6%;  Pred. No. 0.02;
  Matches  145;  Conservative   98;  Mismatches  265;  Indels  315;  Gaps   34;

Qy         10 NSPVEEIINRLNNDY-----------NLGLQCVADTTLTPH-----------------RR 41

Db        381 NEELEALENDVSNGWDPNDMFRYNEENYGVVSTYDSSLSSYTVPLERDNSEEFLKREARA 440

Qy         42 KELAESDEDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAA---------SSNWV 91

Db        441 NQLAEEIESSAQ----YKARVAL---ENDDRSEEEKYTAVQRNSSEREGHSINTRENKYI 493

Qy         92 PKAHADPDTLPWS---KEPPRAATAG--------------------QQWALQTVLLEVL 127

Db        494 PPGQRNREVISWGSGRQNSPRMGQPGSGSMPSRSTSHTSDFNPNSGSDQ--------RVV 545

Qy        128 NRFMPPPNNTP---GRTFGRTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLTRSAT 184

Db        546 NGGVPWPSPCPSPSSRPPSRYQSGPNSLPPRAATPTR---PPSR-----PPSRP-SRPPS 596

Qy        185 GPPIHGAAIPLKFPDPVNTGSKR----------PSLESENLNQCTKRAKGKLSDNV---- 230

Db        597 HPSAHGS------PAPVSTMPKRMSSEGPPRMSPKAQRHPRNHRVSAGRGSISSGLEFVS 650

Qy        231 ----AAAAAPPVPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFS 286

Db        651 HNPPSEAATPPVARTS------------------------------PSGGT--WSSVVS 677

Qy        287 A-------CRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQ 339

Db        678 GVPRLSPKTHRPRSPRQNSIGNTPSGPVLASPQAGIIPTEAVAMPIPAASPTPASPASNR 737

Qy        340 PLSSSSQGETS-FSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARL 398

Db        738 AVTPSSEAKDSRLQDQRQNSPAGNKENIKPNET-SPSFSKAENKGIS------PVVSEHR 790

Qy        399 RNI--WPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKT 456

Db        791 KQIDDLKKFKNDFRLQPSSTSESMDQL-------------LNKNREGEKSRDLI----- 831

Qy        457 LYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLV--- 513

Db        832 ---------KDKIEPSAKDSFIENSSSNCTSGSS----------KPNSPSISPSILSNTE 872

Qy        514 -KLKPLMFEQGCRLT-----------------------------RRFGPDRFFEILIPS 542

Db        873 HKRGPEVTSQGVQTSSPACKQEKDDKEEKKDAAEQVRKSTLNPNAKEFNPRSFSQ---PK 929

Qy        543 PTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAE 602

Db        930 PSTTPTSPRPQAQPSP--------------SMVGHQ--------------------- 951

Qy        603 DPKPIIKERVHFFAETGITPRPDVFKTRSVVPAEEPVE--QRTEFKVSQMLDWLLQLDNN 660

Db        952 QPTPVYTQPV--------CFAPNMMYPVPVSPGVQPLYPIPMTPMPVNQ----------- 992

Qy        661 TWQPHLKLFSRIQLGLSKTYAIMTLEP--HQIRHHKTDLLSPS 701

Db        993 ---------------AKTYRAVPNMPQQRQDQHHQSAMMHPA 1019
```

RESULT 8
US-09-648-281-2
; Sequence 2, Application US/09648281
; Patent No. 6515197
; GENERAL INFORMATION:
; APPLICANT: Pulst, Stefan M.
; TITLE OF INVENTION: Transgenic Animal Model of
; TITLE OF INVENTION: Neurodegenerative Disease and Methods of Use
; FILE REFERENCE: P-CE 4336
; CURRENT APPLICATION NUMBER: US/09/648,281
; CURRENT FILING DATE: 2000-08-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1312
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-648-281-2

```
  Query Match              1.9%;  Score 139.5;  DB 4;  Length 1312;
  Best Local Similarity   17.6%;  Pred. No. 0.02;
  Matches  145;  Conservative   98;  Mismatches  265;  Indels  315;  Gaps   34;

Qy         10 NSPVEEIINRLNNDY-----------NLGLQCVADTTLTPH-----------------RR 41

Db        381 NEELEALENDVSNGWDPNDMFRYNEENYGVVSTYDSSLSSYTVPLERDNSEEFLKREARA 440

Qy         42 KELAESDEDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAA---------SSNWV 91

Db        441 NQLAEEIESSAQ----YKARVAL---ENDDRSEEEKYTAVQRNSSEREGHSINTRENKYI 493

Qy         92 PKAHADPDTLPWS---KEPPRAATAG--------------------QQWALQTVLLEVL 127

Db        494 PPGQRNREVISWGSGRQNSPRMGQPGSGSMPSRSTSHTSDFNPNSGSDQ--------RVV 545

Qy        128 NRFMPPPNNTP---GRTFGRTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLTRSAT 184
```

```
Db         546 NGGVPWPSPCPSPSSRPPSRYQSGPNSLPPRAATPTR---PPSR-----PPSRP-SRPPS 596

Qy         185 GPPIHGAAIPLKFPDPVNTGSKR----------PSLESENLNQCTKRAKGKLSDNV---- 230

Db         597 HPSAHGS------PAPVSTMPKRMSSEGPPRMSPKAQRHPRNHRVSAGRGSISSGLEFVS 650

Qy         231 ----AAAAAPPVPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFS 286

Db         651 HNPPSEAATPPVARTS------------------------------PSGGT--WSSVVS 677

Qy         287 A-------CRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQ 339

Db         678 GVPRLSPKTHRPRSPRQNSIGNTPSGPVLASPQAGIIPTEAVAMPIPAASPTPASPASNR 737

Qy         340 PLSSSSQGETS-FSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARL 398

Db         738 AVTPSSEAKDSRLQDQRQNSPAGNKENIKPNET-SPSFSKAENKGIS------PVVSEHR 790

Qy         399 RNI--WPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKT 456

Db         791 KQIDDLKKFKNDFRLQPSSTSESMDQL--------------LNKNREGEKSRDLI----- 831

Qy         457 LYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLV--- 513

Db         832 ---------KDKIEPSAKDSFIENSSSNCTSGSS----------KPNSPSISPSILSNTE 872

Qy         514 -KLKPLMFEQGCRLT-----------------------------RRFGPDRFFEILIPS 542

Db         873 HKRGPEVTSQGVQTSSPACKQEKDDKEEKKDAAEQVRKSTLNPNAKEFNPRSFSQ---PK 929

Qy         543 PTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAE 602

Db         930 PSTTPTSPRPQAQPSP---------------SMVGHQ---------------------- 951

Qy         603 DPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVE--QRTEFKVSQMLDWLLQLDNN 660

Db         952 QPTPVYTQPV--------CFAPNMMYPVPVSPGVQPLYPIPMTPMPVNQ----------- 992

Qy         661 TWQPHLKLFSRIQLGLSKTYAIMTLEP--HQIRHHKTDLLSPS 701

Db         993 ----------------AKTYRAVPNMPQQRQDQHHQSAMMHPA 1019
```

```
RESULT 9
US-09-707-919A-19
; Sequence 19, Application US/09707919A
; Patent No. 6623927
; GENERAL INFORMATION:
;  APPLICANT: Council of Scientific and Industrial Research
;  TITLE OF INVENTION: Method for detection of human spinocerebeller ataxia 2
;  TITLE OF INVENTION:  gene variants
;  FILE REFERENCE: US 443
;  CURRENT APPLICATION NUMBER: US/09/707,919A
;  CURRENT FILING DATE:  2003-02-10
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
;   LENGTH: 1312
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-707-919A-19

  Query Match            1.9%;  Score 139.5;  DB 4;  Length 1312;
  Best Local Similarity  17.6%;  Pred. No. 0.02;
  Matches  145;  Conservative   98;  Mismatches  265;  Indels  315;  Gaps   34;

Qy          10 NSPVEEIINRLNNDY-----------NLGLQCVADTTLTPH-----------------RR 41

Db         381 NEELEALENDVSNGWDPNDMFRYNEENYGVVSTYDSSLSSYTVPLERDNSEEFLKREARA 440

Qy          42 KELAESDEDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAA----------SSNWV 91

Db         441 NQLAEEIESSAQ----YKARVAL---ENDDRSEEEKYTAVQRNSSEREGHSINTRENKYI 493

Qy          92 PKAHADPDTLPWS---KEPPRAATAG--------------------QQWALQTVLLEVL 127

Db         494 PPGQRNREVISWGSGRQNSPRMGQPGSGSMPSRSTSHTSDFNPNSGSDQ--------RVV 545

Qy         128 NRFMPPPNNTP---GRTFGRTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLTRSAT 184

Db         546 NGGVPWPSPCPSPSSRPPSRYQSGPNSLPPRAATPTR---PPSR-----PPSRP-SRPPS 596

Qy         185 GPPIHGAAIPLKFPDPVNTGSKR----------PSLESENLNQCTKRAKGKLSDNV---- 230

Db         597 HPSAHGS------PAPVSTMPKRMSSEGPPRMSPKAQRHPRNHRVSAGRGSISSGLEFVS 650

Qy         231 ----AAAAAPPVPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFS 286

Db         651 HNPPSEAATPPVARTS------------------------------PSGGT--WSSVVS 677

Qy         287 A-------CRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQ 339

Db         678 GVPRLSPKTHRPRSPRQNSIGNTPSGPVLASPQAGIIPTEAVAMPIPAASPTPASPASNR 737

Qy         340 PLSSSSQGETS-FSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARL 398

Db         738 AVTPSSEAKDSRLQDQRQNSPAGNKENIKPNET-SPSFSKAENKGIS------PVVSEHR 790

Qy         399 RNI--WPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKT 456

Db         791 KQIDDLKKFKNDFRLQPSSTSESMDQL--------------LNKNREGEKSRDLI----- 831

Qy         457 LYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLV--- 513

Db         832 ---------KDKIEPSAKDSFIENSSSNCTSGSS----------KPNSPSISPSILSNTE 872

Qy         514 -KLKPLMFEQGCRLT-----------------------------RRFGPDRFFEILIPS 542

Db         873 HKRGPEVTSQGVQTSSPACKQEKDDKEEKKDAAEQVRKSTLNPNAKEFNPRSFSQ---PK 929

Qy         543 PTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAE 602

Db         930 PSTTPTSPRPQAQPSP---------------SMVGHQ---------------------- 951

Qy         603 DPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVE--QRTEFKVSQMLDWLLQLDNN 660

Db         952 QPTPVYTQPV--------CFAPNMMYPVPVSPGVQPLYPIPMTPMPVNQ----------- 992

Qy         661 TWQPHLKLFSRIQLGLSKTYAIMTLEP--HQIRHHKTDLLSPS 701

Db         993 ----------------AKTYRAVPNMPQQRQDQHHQSAMMHPA 1019
```

```
RESULT 10
US-09-083-268-3
; Sequence 3, Application US/09083268
; Patent No. 6673535
;  GENERAL INFORMATION:
;    APPLICANT:  Pulst, Stefan M
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING SPINOCEREBELLAR
;    TITLE OF INVENTION:  ATAXIA-2 AND PRODUCTS RELATED THERETO
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mueting, Raasch & Gebhardt, P.A.
;      STREET:  119 No. 6673535th Fourth Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  USA
;      ZIP:  55401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/09/083,268
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/727,084
;      FILING DATE:  08-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McCormack, Myra H
;      REGISTRATION NUMBER:  36,602
;      REFERENCE/DOCKET NUMBER:  232.00010101
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612/305-1220
;      TELEFAX:  612/305-1228
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1312 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-083-268-3

  Query Match             1.9%;  Score 139.5;  DB 4;  Length 1312;
  Best Local Similarity   17.6%;  Pred. No. 0.02;
  Matches  145;  Conservative   98;  Mismatches  265;  Indels  315;  Gaps   34;

Qy         10 NSPVEEIINRLNNDY----------NLGLQCVADTTLTPH----------------RR 41
Db        381 NEELEALENDVSNGWDPNDMFRYNEENYGVVSTYDSSLSSYTVPLERDNSEEFLKREARA 440

Qy         42 KELAESDEDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAA----------SSNWV 91
Db        441 NQLAEEIESSAQ----YKARVAL---ENDDRSEEEKYTAVQRNSSEREGHSINTRENKYI 493

Qy         92 PKAHADPDTLPWS---KEPPRAATAG--------------------QQWALQTVLLEVL 127
Db        494 PPGQRNREVISWGSGRQNSPRMGQPGSGSMPSRSTSHTSDFNPNSGSDQ--------RVV 545

Qy        128 NRFMPPPNNTP---GRTFGRTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLTRSAT 184
Db        546 NGGVPWPSPCPSPSSRPPSRYQSGPNSLPPRAATPTR---PPSR-----PPSRP-SRPPS 596

Qy        185 GPPIHGAAIPLKFPDPVNTGSKR----------PSLESENLNQCTKRAKGKLSDNV---- 230
Db        597 HPSAHGS------PAPVSTMPKRMSSEGPPRMSPKAQRHPRNHRVSAGRGSISSGLEFVS 650

Qy        231 ----AAAAAPPVPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFS 286
Db        651 HNPPSEAATPPVARTS------------------------------PSGGT--WSSVVS 677

Qy        287 A-------CRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQ 339
Db        678 GVPRLSPKTHRPRSPRQNSIGNTPSGPVLASPQAGIIPTEAVAMPIPAASPTPASPASNR 737

Qy        340 PLSSSSQGETS-FSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARL 398
Db        738 AVTPSSEAKDSRLQDQRQNSPAGNKENIKPNET-SPSFSKAENKGIS------PVVSEHR 790

Qy        399 RNI--WPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKT 456
Db        791 KQIDDLKKFKNDFRLQPSSTSESMDQL--------------LNKNREGEKSRDLI----- 831

Qy        457 LYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLV--- 513
Db        832 ---------KDKIEPSAKDSFIENSSSNCTSGSS----------KPNSPSISPSILSNTE 872

Qy        514 -KLKPLMFEQGCRLT-----------------------------RRFGPDRFFEILIPS 542
Db        873 HKRGPEVTSQGVQTSSPACKQEKDDKEEKKDAAEQVRKSTLNPNAKEFNPRSFSQ---PK 929

Qy        543 PTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAE 602
Db        930 PSTTPTSPRPQAQPSP---------------SMVGHQ---------------------- 951

Qy        603 DPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVE--QRTEFKVSQMLDWLLQLDNN 660
Db        952 QPTPVYTQPV--------CFAPNMMYPVPVSPGVQPLYPIPMTPMPVNQ----------- 992

Qy        661 TWQPHLKLFSRIQLGLSKTYAIMTLEP--HQIRHHKTDLLSPS 701
Db        993 ----------------AKTYRAVPNMPQQRQDQHHQSAMMHPA 1019

RESULT 11
US-08-976-255-11
; Sequence 11, Application US/08976255
; Patent No. 6136581
;  GENERAL INFORMATION:
;    APPLICANT:  Jono, Keith E.
;    APPLICANT:  Plowman, Gregory
;    TITLE OF INVENTION:  KINASE GENES AND USES
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/976,255
;      FILING DATE:  No. 6136581ember 21, 1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/031,675
;      FILING DATE:  No. 6136581ember 22, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  229/182
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1384 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Protein
US-08-976-255-11

  Query Match             1.9%;  Score 139.5;  DB 3;  Length 1384;
  Best Local Similarity   21.1%;  Pred. No. 0.023;
  Matches  215;  Conservative  114;  Mismatches  371;  Indels  319;  Gaps   57;

Qy         13 VEEIINRLNNDYNLGLQCVAD---TTLTPHRRKELAESDEDFGRHDKIYRALNFLYWRKD 69
Db        478 VTETSRGLNFEYKWEAGRGAEAFPATLSPGRTARLQELCAPDGAPPGVVPVLS----AHS 533

Qy         70 DSLNQAEANFFIEAKAASSNWVPKAHADPDTLPWSKEPPRAATAGQQWALQTVLLEVLNR 129
Db        534 PSLG---SEYFIRLEEAA----PAAGHDPDCAGCAPSPP--ATADQD------------ 571

Qy        130 FMPPPNNTPGRTFGRTLSGP-SGLSRPTSTNTKRKDE-PANVTFADP--PKRSLTRSATG 185
Db        572 -----DDSDGSTAASLAMEPLLGHGPPVDVPWGRGDHYPRRSLARDPLCPSRSPSPSA-G 625

Qy        186 P--PIHGAA----------IPLKFPDPVNTGSKRPSLESENLNQCTKRAKGKLSDNVAAA 233
```

```
Db    626 PLSLAEGGAEDADWGVAAFCPAFFEDPLGTSP------------------------LGSS 661

Qy    234 AAPPVPIA--SALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSS---VFSAC 288

Db    662 GAPPLPLTGEDELEEVGARRAAQ------RGHWRSNV--SANNNSGSRCPESWDPVSAGC 713

Qy    289 RHNQSTTQSSFEAPPSQPREKRPVD-------ATVFEAG------HL----------IE 324

Db    714 --HAEGCPSPKQTPRASPEPGYPGEPLLGLQAASAQEPGCCPGLPHLCSAQGLAPAPCLV 771

Qy    325 SPSKGRTTKS---HIDNQP-LSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSE 380

Db    772 TPSWTETASSGGDHPQAEPKLATEAEGTTGPRLPLPSVPSPSQEGA-PLPS-------EE 823

Qy    381 ESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPL-AVAWEVTRLFMHCKV---------- 429

Db    824 ASA-------------------PDAPDALPDSPTPATGGEVSAIKLASALNGSSSSPEVE 864

Qy    430 -------DLEDESLGLKYDPS----WSTARDVTDIWKTLYRL----DAFRGKPFPEKPPN 474

Db    865 APSSEDEDTAEATSGIFTDTSSDGLQARRPDVVPAFRSLQKQVGTPDSLDSLDIPSSASD 924

Qy    475 ---DVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFG 531

Db    925 GGYEVFSPSATG--PSGGQPRALDS--GYDTENY-ESPEFVLK----EAQEGCE------ 969

Qy    532 PDRFFEILI----PSP-TSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAK 586

Db    970 PQAFAELASEGEGPGPETRLSTSLSGLNEKNPYRDSAYFSDLEAEAEATSGPE-----KK 1024

Qy    587 DAGYRKPLREFQLRAEDPKPIIKERVHFFAETG-----ITFRPDVF-KTRSVVPAEEPVE 640

Db   1025 CGGDRAPGPELGL----------------PSTGQPSEQVCLRPGVSGEAQGSGPGE---- 1064

Qy    641 QRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGL--------------------SKTY 680

Db   1065 ---------VLPPLLQLEGSSPEP-----STCPSGLVPEPPEPQGPAKVRPGPSPSCSQF 1110

Qy    681 AIMTLEPHQIRHHKTDLLSPSG--TGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGR 738

Db   1111 FLLTPVPLRSEGNSSEFQGPPGLLSGPAPQK---RMGGPGTPRAPLRLALPGLPAALEGR 1167

Qy    739 FGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFV-----DKHQRTLEVRSVASELKSAG 793

Db   1168 ------------PEEEEEDSEDSDESDEELRCYSVQEPSEDSEEEAPAVPVVVAESQSA- 1214

Qy    794 LNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRA 853

Db   1215 RNL---------RSLLKMPSLLSETFCEDLER----KKKAVSFFDDVTVYLFDQES--P 1258

Qy    854 TRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQD---------------IAWD 897

Db   1259 TR-ELGE-PF----PGAKESPPTFLRGSPGSPSAPNRPQQADGSPNGSTAEEGGGFAWD 1311
```

RESULT 12
US-09-845-917A-3
; Sequence 3, Application US/09845917A
; Patent No. 6653529
; GENERAL INFORMATION:
; APPLICANT: Bogaert, Thierry
; APPLICANT: Vandekerckhove, Joel
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTIANING THEM AND THEIR
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR
; FILE REFERENCE: P/14-1
; CURRENT APPLICATION NUMBER: US/09/845,917A
; CURRENT FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1528
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-845-917A-3

```
  Query Match             1.9%;  Score 139;  DB 4;  Length 1528;
  Best Local Similarity   21.3%;  Pred. No. 0.03;
  Matches  156;  Conservative   81;  Mismatches  248;  Indels  246;  Gaps   33;

Qy      1 MNPITPRKRNSP-----VEEIINRLNN-----DY--NLGLQCVADTTLTPHRRKELAESD 48

Db     46 INVIVPINEFSPAFTKRLAKITSNLDGLETCLDYLKNLGLDC-----------SKLTKTD 94

Qy     49 EDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLPWSKEP- 107

Db     95 IDSGNLGAVLQLL-FLLSTYKQKLRQ----------------LKKDQKKLEQLPTSIMPP 137

Qy    108 -------PRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNT 160

Db    138 AVSKLPSPRVATSATASATN------------PNSNFPQMSTSRLQTPQSRISKIDSSKI 185

Qy    161 KRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRP------------ 208

Db    186 GIKPKTSGL---KPPSSSTTSSN------------------NTNSFRPSSRSSGNNNVGS 224

Qy    209 -------SLES-------ENLNQCT------KRAKGKL----------SDNVAAAAAPPV 238

Db    225 TISTSAKSLESSSTYSSISNLNRPTSQLQKPSRPQTQLVRVATTTKIGSSKLAAPKAVST 284

Qy    239 PIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSS 298

Db    285 PKLASVKTIGAKQEPDNSGGGGGGMLKLKLFSSKNPSSSSNSPQPTRKAAAVPQQQTLSK 344

Qy    299 FEAPPS---QPREKRPVDATVFEAGHLIESPSKGRTTK-----------SHIDNQPLSSS 344

Db    345 IAAPVKSGLKPPTSKLGSAT---------SMSKLCTPKVSYRKTDAPIISQQDSKRCSKS 395

Qy    345 SQGETSFSTYYESFP-SSGGEGAIPEPSRSNGLARSEESARS--QVQVHAPVVAARLRNI 401

Db    396 SEEESGYAGFNSTSPTSSSTEGSLSMHSTSSKSSTSDEKSPSSDDLTLNASIVTAIRQPI 455

Qy    402 --WPKFP----KWLHEAP-LAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIW 454

Db    456 AATPVSPNIINKPVEEKPTLAVK-------------------GVK-----STAKK----- 486

Qy    455 KTLYRLDAFRGKPFPEKPPNDVFVT-----------AMTGN--FESKGSAVVLSAVLDYN 501

Db    487 -----------DPPPAVPPRDTQPTIGVVSPIMAHKKLTNDPVISEKPEPEKLQSMSIDT 535

Qy    502 PDNSPTAPL-YLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGA 560

Db    536 TDVPPLPPLKSVVPLKMTSIRQ----------PPTYDVLLKQGKITSP-VKSFGYEQSSA 584

Qy    561 VEEVI----------QWLTMGQHSLVGRQWRAFFAKDAGY------------RKPLREFQ 598

Db    585 SEDSIVAHASAQVTPPTKTSGNHSLERRMGKNKTSESSGYTSDAGVAMCAKMREKLKEYD 644

Qy    599 ---LRAEDPKP 606

Db    645 DMTRRAQNGYP 655
```

RESULT 13
US-09-845-917A-4
; Sequence 4, Application US/09845917A
; Patent No. 6653529
; GENERAL INFORMATION:
; APPLICANT: Bogaert, Thierry
; APPLICANT: Vandekerckhove, Joel
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTIANING THEM AND THEIR
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR

```
; FILE REFERENCE: P/14-1
; CURRENT APPLICATION NUMBER: US/09/845,917A
; CURRENT FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 1583
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
US-09-845-917A-4

  Query Match             1.9%;  Score 139;  DB 4;  Length 1583;
  Best Local Similarity   21.3%;  Pred. No. 0.032;
  Matches  156;  Conservative   81;  Mismatches  248;  Indels  246;  Gaps   33;

Qy    1 MNPITPRKRNSP-----VEEIINRLNN-----DY--NLGLQCVADTTLTPHRRKELAESD 48
Db   46 INVIVPINEFSPAFTKRLAKITSNLDGLETCLDYLKNLGLDC-----------SKLTKTD 94

Qy   49 EDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLPWSKEP- 107
Db   95 IDSGNLGAVLQLL-FLLSTYKQKLRQ----------------LKKDQKKLEQLPTSIMPP 137

Qy  108 -------PRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNT 160
Db  138 AVSKLPSPRVATSATASATN------------PNSNFPQMSTSRLQTPQSRISKIDSSKI 185

Qy  161 KRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRP------------ 208
Db  186 GIKPKTSGL---KPPSSSTTSSN------------------NTNSFRPSSRSSGNNNVGS 224

Qy  209 -------SLES-------ENLNQCT------KRAKGKL----------SDNVAAAAAPPV 238
Db  225 TISTSAKSLESSSTYSSISNLNRPTSQLQKPSRPQTQLVRVATTTKIGSSKLAAPKAVST 284

Qy  239 PIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSS 298
Db  285 PKLASVKTIGAKQEPDNSGGGGGGMLKLKLFSSKNPSSSSNSPQPTRKAAAVPQQQTLSK 344

Qy  299 FEAPPS---QPREKRPVDATVFEAGHLIESPSKGRTTK-----------SHIDNQPLSSS 344
Db  345 IAAPVKSGLKPPTSKLGSAT---------SMSKLCTPKVSYRKTDAPIISQQDSKRCSKS 395

Qy  345 SQGETSFSTYYESFP-SSGGEGAIPEPSRSNGLARSEESARS--QVQVHAPVVAARLRNI 401
Db  396 SEEESGYAGFNSTSPTSSSTEGSLSMHSTSSKSSTSDEKSPSSDDLTLNASIVTAIRQPI 455

Qy  402 --WPKFP----KWLHEAP-LAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIW 454
Db  456 AATPVSPNIINKPVEEKPTLAVK-------------------GVK-----STAKK----- 486

Qy  455 KTLYRLDAFRGKPFPEKPPNDVFVT-----------AMTGN--FESKGSAVVLSAVLDYN 501
Db  487 -----------DPPPAVPPRDTQPTIGVVSPIMAHKKLTNDPVISEKPEPEKLQSMSIDT 535

Qy  502 PDNSPTAPL-YLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGA 560
Db  536 TDVPPLPPLKSVVPLKMTSIRQ----------PPTYDVLLKQGKITSP-VKSFGYEQSSA 584

Qy  561 VEEVI-----------QWLTMGQHSLVGRQWRAFFAKDAGY------------RKPLREFQ 598
Db  585 SEDSIVAHASAQVTPPTKTSGNHSLERRMGKNKTSESSGYTSDAGVAMCAKMREKLKEYD 644

Qy  599 ---LRAEDPKP 606
Db  645 DMTRRAQNGYP 655

RESULT 14
US-09-489-039A-11704
; Sequence 11704, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11704
;   LENGTH: 1768
;   TYPE: PRT
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11704

  Query Match             1.9%;  Score 137;  DB 4;  Length 1768;
  Best Local Similarity   19.4%;  Pred. No. 0.058;
  Matches  196;  Conservative  142;  Mismatches  347;  Indels  326;  Gaps   53;

Qy  407 KWLHEAPLAVAWEVTRLFMHCK------------VDLEDESLGL--KYD--PSWSTARDV 450
Db  778 KSLFLKPYGEGWEALFHFIYQERDDLSSPVRTQVIELIDEWTGLINIHDELPKASEKVGF 837

Qy  451 TDIWKTLYRLDAFRGKPFPEKPPNDVF--VTAMTGNFESKGSAVVLSAVLDYNPDNSPTA 508
Db  838 LSLWLLEKVKDSYRGEGQRKKILNALLKVSTAIKDDF-----------------DELMKQ 880

Qy  509 PLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVI--Q 566
Db  881 DVFTSKIKP----------RRLS----YVDELSSLALIGFNVPMLCKFYPDFVVKLSMHE 926

Qy  567 WLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFRPDV 626
Db  927 WLLEESEE-----------NDYGY----GSYRVDVEESFGLDTER-DFFPASG------- 963

Qy  627 FKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKL-FSRIQLGLSKT-YAIMT 684
Db  964 --------AKGPFKYLLQFKPRLGLDFIIRLCNLTAQKYSESDFSKTQDNKEDTIYAFET 1015

Qy  685 LEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRI------RDVLGLGDVPSAVQGR 738
Db 1016 VA------KQVDI---------TLNDG------TVVKQYASPHLWKGYRGQSTLPYLLQSA 1055

Qy  739 FGSAKGMWVID-VDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLN-- 795
Db 1056 LMALEN-WLVEYVENCGKNNEID-------WIFDYVLRSSNSVMPTSVLSSVATGFPNKV 1107

Qy  796 ----LQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALN-------RPVEFRQW 844
Db 1108 GKAAFPLLKTADLYHLDLIRMTQEMGG---NEM-HFFGLNRDALSKIYLEERREAALRPW 1163

Qy  845 VYESYSSRATRVSHGRVPFLAGLPD---------------SQEETLNFLMNSGFDPKKQK 889
Db 1164 RKESLETLLTRLQ-----FVNELRDDILKIVDELKNEATASNEKSLRYMVHR-VDTRTWK 1217

Qy  890 YLQD-------------IAWDL---------QKRKCDTLKSKLNIRVGRSAYIYMIADFWG 928
Db 1218 VVEDKENDRVLFQSSSELPEDLKLDQQEFNEKHAMDNTVTSLNL--------------WG 1263

Qy  929 VLEENEVHVGFSSKFRDEE--ESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLK 986
Db 1264 ------------KKLFDEQLLEEKYFSSYQDALIATKGL---------LSALQKGEVHNFA 1303

Qy  987 DVII--FSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDK 1044
Db 1304 DMAVGTITTVAAVCVRDVLLELSYE-DKKW-CLEIILESIFMHAD-----NMNGTAAHDK 1356

Qy 1045 TTFKQLMASHGTGSAA------------KEQTTYDMIQKSFHFALQPNFLGMCTNYKERL 1092
Db 1357 TDY------YGSGACAFVLPKLFDLDLDSEQIEY------LKFAL------------ATAL 1393

Qy 1093 CYINNSVS-------------NKPAIILSSLVGNLVDQSK--------------QGIVFNE 1126
```

```
Db       1394 THENLNVSACAAKGVREFLWSRDAELASRCIGGIVEYARLRREDSEVRRFYHLQGAELQA 1453

Qy       1127 A--SWAQL----RRELLGG--ALSLPDPMYKSD-SWLGRGEPTHIIDYLKFSIARPAIDK 1177

Db       1454 ALEKWNDLLTSFRNDLIEGKFKLSVNDISLESHCSWFIH-LPMLMVPYCTKD------EI 1506

Qy       1178 ELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISD-------KSRSS---------A 1221

Db       1507 QIQLVNRLMSFVFDTEYQNHRSSDDEKINHEIKKQIQDCLSEHVVSSRKNNFMPFKDLLV 1566

Qy       1222 LLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPEA 1272

Db       1567 LGCTKAPGFIYSLTLSYHVAVEKEGDYDA----------IWALWCLLAPEA 1607
```

RESULT 15
US-09-579-181-2
; Sequence 2, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chrivia, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2972
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-2

Query Match 1.8%; Score 136.5; DB 4; Length 2972;
Best Local Similarity 18.4%; Pred. No. 0.16;
Matches 196; Conservative 117; Mismatches 364; Indels 389; Gaps 44;

```
Qy         82 EAKAASSNWVPKAHADPDTLPWSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTP-GR 140

Db       1125 EVSASAPGAAPLTISSPLHVPSSLPGPASSP-----------------MPIPNSSPLAS 1166

Qy        141 TFGRTLSGPSGLSRPTSTNTK-------------------RKDEPANVTFADPPKRSLTR 181

Db       1167 PVSSTVSVPLSSSLPISVPTTLPAPASAPLTIPISAPLTVSASGPALLTSVTPPLAPVVP 1226

Qy        182 SATGPP------------------------------------------------------ 187

Db       1227 AAPGPPSLQPSGASPSASALTLGLATAPSLSSSQTPGHPLLLAPTSSHVPGLNSTVAPAC 1286

Qy        188 ----IHGAAIPLKFPDPVNTGSKRPSLESENLNQCTKRAKGKLSDNVAAAAAPPVPI--- 240

Db       1287 SPVLVPASALASPFPSAPNPAPAQASLLAP-----ASSASQALATPLAPMAAPQTAILAP 1341

Qy        241 --ASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSS--------VFSACRH 290

Db       1342 SPAPPLAPLPVLAPSPGAAPVLASSQTPVPVMAPSSTPGTSLASASPVPAPTPVLA---- 1397

Qy        291 NQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQPLS-SSSQGET 349

Db       1398 -PSSTQTMLPAPVPSPLPSPASTQTLALAPAL--APTLGGSSPS----QTLSLGTGNPQG 1450

Qy        350 SFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWL 409

Db       1451 PFPT--QTLSLTPASSLVPTPAQTLSLAPGPPLGPTQTLSLAP--APPLAPASPVGPAPA 1506

Qy        410 HEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFP 469

Db       1507 HTLTLAPASSSASLLAPASVQT------LTLSPA----------------------PVP 1537

Qy        470 EKPP----NDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCR 525

Db       1538 TLGPAAAQTLALAPASTQSPASQASSLVVSA------SGAAPLPVTMVSRLPVSKDEPDT 1591

Qy        526 LTRRFGPDRFFEILIPSPTSTSPSV--------PPVVSKQP------------------ 558

Db       1592 LTLRSGP--------PSPPSTATSFGGPRPRRQPPPPPRSPFYLDSLEEKRKRQRSERLE 1643

Qy        559 ---------GAV-----EEVIQWLTMGQ--HSLVGRQ---------WRAFFAKDAGYRKP 593

Db       1644 RIFQLSEAHGALAPVYGTEVLDFCTLPQPVASPIGPRSPGPSHPTFWT---YTEAAHRAV 1700

Qy        594 LREFQ------------------------LRAEDPKPIIKERVHFFAETGITFRPDVFK 628

Db       1701 LFPQQRLDQLSEIIERFIFVMPPVEAPPPSLHACHPPPWLAPRQAAFQEQ---------L 1751

Qy        629 TRSVVPAEEPVEQ-----RTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKTYAIM 683

Db       1752 ASELWPRARPLHRIVCNMRTQF------------------PDLRLI-QYDCGKLQTLAVL 1792

Qy        684 TLEPHQIRHHKT--------------DLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLG 729

Db       1793 -LRQLKAEGHRVLIFTQMTRMLDVLEQFLTYHGHLYLRLDGSTRVEQR------------ 1839

Qy        730 DVPSAVQGRFGSAKGMWV---------IDVDDTG-------DEDWIETYPSQRKWECDFV 773

Db       1840 ---QALMERFNADKRIFCFILSTRSGGVGVNLTGADTVVFYDSDWNPTMDAQAQDRCHRI 1896

Qy        774 DKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLI---NDLQRQFSE 830

Db       1897 GQ-TRDVHIYRLISE-----------RTVEENILKKANQKRMLGDMAIEGGNFTTAYFKQ 1944

Qy        831 Q--KHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQ 888

Db       1945 QTIRELFDMPLE--------------EPSSSSVP---SAPEEEETV---------ASKQ 1978

Qy        889 KYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENE 934

Db       1979 THI------LEQALC-RAEDEEDIRAATQAKAEQVAE---LAEFNE 2014
```

Search completed: March 29, 2004, 08:50:39
Job time : 33 secs

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         March 31, 2004, 02:06:52 ; Search time 8802.43 Seconds
                                           (without alignments)
                                           6903.427 Million cell updates/sec

Title:          US-09-913-878A-2
Perfect score:  7397
Sequence:       1 MNPITPRKRNSPVEEIINRL..........YEVLGDDDFDGIGFTGNGDY 1402

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:      6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09913878/runat_29032004_082008_8366/app_query.fasta_1.2254
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878_@CGN_1_1_7684_@runat_29032004_082008_8366 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	7397	100.0	4206	8	NCR133528	AJ133528 Neurospor
	2	7397	100.0	8045	6	AX034425	AX034425 Sequence
c	3	348.5	4.7	165701	8	AP004357	AP004357 Oryza sat
	4	328	4.4	3505	8	NTA011576	AJ011576 Nicotiana
	5	328	4.4	3591	6	AX204840	AX204840 Sequence
c	6	306	4.1	82584	8	NCB13B3	BX284762 Neurospor
	7	299	4.0	6847	8	AY029284	AY029284 Neurospor
	8	298	4.0	5992	3	DDI314910	AJ314910 Dictyoste
	9	297.5	4.0	3731	6	AR145904	AR145904 Sequence
	10	297.5	4.0	3731	8	LERDRP	Y10403 L.esculentu
	11	296.5	4.0	6688	8	AF468822	AF468822 Diaporthe
	12	293.5	4.0	5731	3	DDI314909	AJ314909 Dictyoste
	13	290	3.9	3807	8	AY148431	AY148431 Arabidops
	14	289.5	3.9	4013	8	AF239718	AF239718 Arabidops
	15	289.5	3.9	6863	6	AX204839	AX204839 Sequence
c	16	289.5	3.9	80367	8	AC012329	AC012329 Arabidops
c	17	289.5	3.9	104204	8	ATT9C5	AL132964 Arabidops
	18	285.5	3.9	1538	8	ATH011977	AJ011977 Arabidops
	19	284.5	3.8	5280	8	AF268093	AF268093 Arabidops
	20	281	3.8	6987	8	AY049072	AY049072 Diaporthe
c	21	277.5	3.8	132716	2	AC105931	AC105931 Magnaport
	22	273.5	3.7	3524	8	AF443073	AF443073 Phomopsis
	23	268	3.6	44658	8	SPAC6F12	Z98533 S.pombe chr
	24	265	3.6	7178	8	AF500110	AF500110 Neurospor
	25	261.5	3.5	29132	3	CEF10B5	Z48334 Caenorhabdi
	26	261	3.5	6204	8	AF411016	AF411016 Neurospor
	27	253.5	3.4	2639	8	AK101440	AK101440 Oryza sat
c	28	247	3.3	132699	8	AC006917	AC006917 Genomic s
	29	242.5	3.3	121337	8	OSJN00104	AL606653 Oryza sat
	30	236.5	3.2	4978	3	AF117611	AF117611 Dictyoste
	31	236	3.2	5025	3	AF159144	AF159144 Caenorhab
	32	230	3.1	5208	3	AF159143	AF159143 Caenorhab
c	33	225	3.0	74406	8	F2P3	AF080120 Arabidops
	34	225	3.0	108598	8	ATT22B4	AL049876 Arabidops
	35	225	3.0	198301	8	ATCHRIV31	AL161531 Arabidops
	36	218.5	3.0	4300	3	AF293414	AF293414 Giardia i
	37	217.5	2.9	6103	3	AF393743	AF393743 Giardia i
	38	211.5	2.9	299925	1	AP005045	AP005045 Streptomy
	39	209.5	2.8	3628	8	AK110169	AK110169 Oryza sat
	40	199.5	2.7	4579	8	PHY011979	AJ011979 Petunia h
	41	196	2.6	308050	1	SCO939124	AL939124 Streptomy
c	42	195.5	2.6	17589	1	AE004862	AE004862 Pseudomon
c	43	195	2.6	11908	1	AE005822	AE005822 Caulobact
c	44	194.5	2.6	31521	3	CEY47H10A	AL132848 Caenorhab
	45	192.5	2.6	300350	1	BX294145	BX294145 Pirellula

ALIGNMENTS

RESULT 1

```
NCR133528
LOCUS       NCR133528               4206 bp    DNA     linear   PLN 10-MAY-1999
DEFINITION  Neurospora crassa qde-1 gene, partial.
ACCESSION   AJ133528
VERSION     AJ133528.1  GI:4803726
KEYWORDS    qde-1 gene; RNA-dependent RNA polymerase.
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Cogoni,C. and Macino,G.
  TITLE     Gene silencing in neurospora crassa requires a protein homologous
            to RNA-dependent RNA polymerase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4206)
  AUTHORS   Cogoni,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-1999) Cogoni C., Dept. Biotecnologie Cellulari ed
            Ematologia,, Universita di Roma La Sapienza, Viale Regina Elena
            324, 00161 Rome, ITALY
FEATURES             Location/Qualifiers
     source          1..4206
                     /organism="Neurospora crassa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5141"
     gene            1..4206
                     /gene="qde-1"
     CDS             1..>4206
                     /gene="qde-1"
                     /function="gene silencing"
                     /note="putative"
                     /codon_start=1
                     /product="RNA-dependent RNA polymerase"
                     /protein_id="CAB42634.1"
                     /db_xref="GI:4803727"
                     /db_xref="GOA:Q9Y7G6"
                     /db_xref="SPTREMBL:Q9Y7G6"
                     /translation="MNPITPRKRNSPVEEIINRLNNDYNLGLQCVADTTLTPHRRKEL
                     AESDEDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLP
                     WSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNT
                     KRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQC
                     TKRAKGKLSDNVAAAAAPPVPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQ
                     GTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKS
                     HIDNQPLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAP
                     VVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDV
                     TDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTA
                     PLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQ
                     WLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFRP
                     DVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKTYAI
                     MTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFG
                     SAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQL
                     LPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRV
                     SHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRV
                     GRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQ
                     RVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEM
                     PLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNY
                     KERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDP
                     MYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWDPDLA
                     SYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQVY
                     EKWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHKSPKF
                     VWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGDGSEYP
                     DPEVYEVLGDDDFDGIGFTGNGDY"
ORIGIN

Alignment Scores:
Pred. No.:               0           Length:         4206
Score:                   7397.00     Matches:        1402
Percent Similarity:      100.00%     Conservative:   0
Best Local Similarity:   100.00%     Mismatches:     0
Query Match:             100.00%     Indels:         0
DB:                      8           Gaps:           0
```

```
US-09-913-878A-2 (1-1402) x NCR133528 (1-4206)

Qy        1 MetAsnProIleThrProArgLysArgAsnSerProValGluGluIleIleAsnArgLeu 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 ATGAACCCTATTACTCCTAGGAAGAGGAATAGCCCCGTCGAGGAAATCATAAACCGGCTC 60

Qy       21 AsnAsnAspTyrAsnLeuGlyLeuGlnCysValAlaAspThrThrLeuThrProHisArg 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 AATAACGACTACAACCTGGGCCTCCAGTGTGTCGCAGACACAACTCTCACCCCCCACCGC 120

Qy       41 ArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHisAspLysIleTyrArgAla 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 CGGAAGGAGCTGGCCGAGAGTGACGAGGATTTCGGTCGCCATGACAAGATCTACAGAGCC 180

Qy       61 LeuAsnPheLeuTyrTrpArgLysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePhe 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 CTGAACTTTCTCTACTGGCGGAAGGATGACTCCCTGAACCAGGCAGAAGCCAACTTCTTC 240

Qy       81 IleGluAlaLysAlaAlaSerSerAsnTrpValProLysAlaHisAlaAspProAspThr 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 ATCGAGGCCAAAGCTGCGAGCTCGAACTGGGTGCCCAAAGCCCACGCCGACCCTGACACG 300

Qy      101 LeuProTrpSerLysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGln 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 CTTCCGTGGTCCAAGGAACCTCCCCGCGCCGCTACTGCCGGCCAACAATGGGCATTGCAG 360

Qy      121 ThrValLeuLeuGluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArg 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 ACTGTGTTGCTCGAGGTGCTTAATAGGTTTATGCCACCTCCCAATAACACACCAGGTCGA 420

Qy      141 ThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThr 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 ACGTTTGGCAGAACTCTAAGCGGCCCAAGTGGCCTGAGCCGCCCAACCTCTACCAACACC 480

Qy      161 LysArgLysAspGluProAlaAsnValThrPheAlaAspProProLysArgSerLeuThr 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 AAACGCAAGGATGAGCCCGCCAATGTCACTTTCGCTGATCCGCCCAAACGCTCGTTGACT 540

Qy      181 ArgSerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspPro 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 CGCTCTGCCACAGGTCCTCCTATTCACGGCGCGGCGATACCCCTAAAGTTCCCCGATCCA 600

Qy      201 ValAsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLys 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 GTGAATACCGGTTCCAAACGACCATCTCTCGAGAGTGAGAATCTCAATCAGTGCACCAAG 660

Qy      221 ArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValProIle 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 CGGGCCAAGGGCAAGCTGTCTGATAATGTTGCCGCTGCCGCCGCCCCGCCCGTGCCTATT 720

Qy      241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 GCGAGCGCTTTGGACAAGGTACCGACTCGAAGGCATGCCAATACGAGAGATCCCACGGCG 780

Qy      261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 ACAGGTCATAGACGAGCGGACCAGGTGGATTCCTTTGATACATCTCAAGGCACTTCCTAT 840

Qy      281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 GGTTCGAGTGTCTTCAGCGCTTGCCGTCACAATCAGAGCACTACCCAGAGTAGTTTTGAG 900

Qy      301 AlaProProSerGlnProArgGluLysArgProValAspAlaThrValPheGluAlaGly 320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 GCTCCTCCTTCACAGCCCAGAGAGAAGCGGCCTGTGGATGCCACGGTCTTTGAGGCTGGA 960

Qy      321 HisLeuIleGluSerProSerLysGlyArgThrThrLysSerHisIleAspAsnGlnPro 340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 CACTTGATTGAGTCTCCTAGCAAAGGAAGAACAACCAAGTCCCACATAGATAACCAGCCC 1020

Qy      341 LeuSerSerSerSerGlnGlyGluThrSerPheSerThrTyrTyrGluSerPheProSer 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 CTTTCATCGTCTTCCCAGGGTGAAACTTCGTTCAGCACTTACTATGAGTCGTTTCCAAGT 1080
```

```
Qy    361 SerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLeuAlaArgSerGlu 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCCGGCGGCGAGGGCGCAATTCCCGAGCCGAGTCGCTCAAATGGACTGGCTCGGAGCGAA 1140

Qy    381 GluSerAlaArgSerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsn 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GAAAGCGCTCGATCTCAGGTTCAAGTTCATGCCCCGGTGGTTGCAGCTCGGCTGAGAAAT 1200

Qy    401 IleTrpProLysPheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluVal 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ATTTGGCCGAAATTTCCCAAATGGCTACACGAAGCTCCTCTCGCTGTTGCATGGGAAGTT 1260

Qy    421 ThrArgLeuPheMetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyr 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ACCAGACTCTTTATGCACTGCAAAGTAGACTTGGAAGACGAGAGCCTGGGCCTAAAGTAC 1320

Qy    441 AspProSerTrpSerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeu 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GACCCTTCCTGGTCTACCGCGCGCGATGTCACAGATATCTGGAAGACTCTCTACCGGCTT 1380

Qy    461 AspAlaPheArgGlyLysProPheProGluLysProProAsnAspValPheValThrAla 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GATGCTTTCCGTGGTAAACCCTTTCCAGAAAAGCCGCCCAACGACGTGTTCGTGACCGCA 1440

Qy    481 MetThrGlyAsnPheGluSerLysGlySerAlaValValLeuSerAlaValLeuAspTyr 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ATGACGGGCAACTTTGAGAGCAAAGGTAGTGCCGTTGTTCTCTCTGCTGTTCTAGACTAC 1500

Qy    501 AsnProAspAsnSerProThrAlaProLeuTyrLeuValLysLeuLysProLeuMetPhe 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AATCCGGACAACTCGCCTACTGCGCCCCTTTACCTTGTGAAGCTGAAGCCGCTCATGTTC 1560

Qy    521 GluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIleLeuIle 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GAGCAGGGCTGTCGACTCACCCGTCGGTTCGGTCCTGATAGGTTTTTCGAGATCCTTATA 1620

Qy    541 ProSerProThrSerThrSerProSerValProProValValSerLysGlnProGlyAla 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CCGTCGCCTACGAGCACCAGCCCAAGTGTACCGCCGGTGGTCAGCAAACAACCAGGTGCG 1680

Qy    561 ValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrp 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GTCGAAGAAGTCATCCAGTGGCTCACGATGGGGCAACATTCTCTTGTAGGCCGCCAATGG 1740

Qy    581 ArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CGCGCTTTCTTCGCCAAAGATGCCGGATACAGGAAACCTCTCAGGGAGTTCCAGCTCCGC 1800

Qy    601 AlaGluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIle 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 GCCGAGGACCCGAAACCCATCATCAAGGAGAGAGTCCACTTCTTTGCCGAGACCGGCATT 1860

Qy    621 ThrPheArgProAspValPheLysThrArgSerValValProAlaGluGluProValGlu 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ACGTTCCGACCTGATGTGTTCAAGACGAGATCTGTCGTTCCGGCAGAGGAACCTGTAGAG 1920

Qy    641 GlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAspAsnAsn 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CAACGGACCGAGTTCAAAGTTAGTCAAATGCTGGACTGGCTCCTGCAACTCGACAACAAC 1980

Qy    661 ThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTyr 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 ACTTGGCAGCCGCACCTCAAGTTGTTCTCCCGTATCCAGCTCGGTCTGAGTAAGACATAT 2040

Qy    681 AlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAspLeuLeuSerPro 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 GCCATTATGACATTGGAGCCTCACCAGATCAGACACCACAAGACCGATCTTCTTTCACCT 2100

Qy    701 SerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSerValAlaLys 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TCAGGCACTGGCGAAGTGATGAATGACGGTGTAGGCCGCATGTCGCGAAGCGTGGCCAAG 2160

Qy    721 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGly 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AGGATACGCGATGTTCTCGGTTTGGGTGATGTGCCCTCTGCTGTGCAAGGGCGGTTTGGT 2220

Qy    741 SerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGlu 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TCGGCCAAGGGAATGTGGGTTATCGACGTTGACGACACAGGCGATGAGGATTGGATCGAG 2280

Qy    761 ThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 ACATACCCGTCCCAGCGCAAGTGGGAATGCGACTTCGTTGATAAACATCAACGTACCCTC 2340

Qy    781 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuPro 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GAAGTCCGGAGCGTGGCTTCTGAACTGAAGTCAGCTGGTCTCAACCTACAGCTGTTACCT 2400

Qy    801 ValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeu 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GTCCTGGAAGATAGAGCCAGGGACAAGGTGAAGATGCGCCAGGCAATCGGTGACCGTCTT 2460

Qy    821 IleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 ATCAACGATTTGCAACGACAGTTCAGCGAGCAAAAGCATGCTTTGAATCGCCCAGTGGAA 2520

Qy    841 PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArg 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 TTTCGCCAATGGGTTTACGAGAGTTATTCCAGTCGCGCAACTCGAGTCAGCCACGGCCGT 2580

Qy    861 ValProPheLeuAlaGlyLeuProAspSerGlnGluThrLeuAsnPheLeuMetAsn 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 GTGCCTTTTCTTGCTGGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAAC 2640

Qy    881 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 AGTGGGTTCGATCCCAAGAAGCAAAAGTACTTGCAAGACATCGCCTGGGATCTTCAAAAG 2700

Qy    901 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CGGAAATGTGACACGTTGAAGTCCAAGCTGAACATCCGTGTCGGTCGATCAGCATACATT 2760

Qy    921 TyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TACATGATTGCCGATTTCTGGGGTGTGCTTGAGGAAAATGAGGTTCATGTCGGATTCTCC 2820

Qy    941 SerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 TCAAAGTTCAGAGACGAGGAGGAGTCTTTTACACTCCTATCGGACTGTGATGTCCTCGTG 2880

Qy    961 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 GCGCGATCCCCAGCCCATTTCCCTAGTGATATCCAACGGGTTCGAGCAGTCTTCAAGCCA 2940

Qy    981 GluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAla 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 GAGCTCCACAGTCTCAAGGATGTAATCATCTTCTCTACTAAAGGAGATGTACCGCTTGCT 3000

Qy   1001 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 AAGAAGCTATCTGGTGGAGACTACGACGGCGATATGGCCTGGGTCTGCTGGGATCCGGAG 3060

Qy   1021 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATCGTCGATGGTTTCGTCAATGCGGAAATGCCTCTGGAGCCCGACCTGTCTAGGTACCTA 3120

Qy   1041 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AAGAAGGACAAAACGACTTTCAAACAACTTATGGCCTCACACGGCACGGGCTCAGCGGCC 3180

Qy   1061 LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AAAGAGCAGACTACATACGATATGATCCAGAAGAGCTTCCATTTCGCCCTGCAGCCCAAC 3240

Qy   1081 PheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSer 1100
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 TTCTTGGGCATGTGCACTAACTACAAAGAAAGGCTCTGTTACATCAACAATAGTGTGTCT 3300

Qy       1101 AsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGln 1120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 AACAAGCCGGCCATCATTCTTAGTTCACTGGTGGGAAACCTCGTCGATCAGAGCAAGCAA 3360

Qy       1121 GlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAla 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 GGTATTGTCTTTAACGAAGCAAGCTGGGCTCAATTGCGTAGGGAACTGCTTGGCGGTGCA 3420

Qy       1141 LeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThr 1160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 TTGTCCCTTCCTGACCCAATGTACAAGAGCGACAGTTGGCTCGGGCGCGGAGAGCCTACC 3480

Qy       1161 HisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGlu 1180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 CACATTATTGACTACCTGAAATTCTCCATCGCCAGGCCTGCGATTGACAAGGAACTGGAA 3540

Qy       1181 AlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAsp 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 GCCTTCCACAATGCCATGAAAGCGGCCAAGGATACAGAAGACGGCGCTCACTTTTGGGAT 3600

Qy       1201 ProAspLeuAlaSerTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSer 1220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCGGATCTCGCTTCCTACTACACGTTCTTCAAGGAGATTAGCGACAAGTCGCGATCGTCC 3660

Qy       1221 AlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArg 1240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 GCACTGCTATTCACGACTCTGAAGAACCGTATCGGCGAAGTCGAGAAAGAATATGGCAGG 3720

Qy       1241 LeuValLysAsnLysGluMetArgAspSerLysAspProTyrProValArgValAsnGln 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 TTGGTCAAAAACAAGGAGATGAGAGACAGCAAGGACCCCTACCCTGTCCGCGTCAACCAG 3780

Qy       1261 ValTyrGluLysTrpCysAlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyr 1280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 GTTTATGAAAAATGGTGCGCCATCACGCCTGAGGCGATGGACAAATCCGGAGCAAATTAT 3840

Qy       1281 AspSerLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThr 1300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 GATTCTAAGGTGATCAGGTTGCTGGAGCTGTCCTTCCTCGCGGACCGTGAGATGAATACA 3900

Qy       1301 TrpAlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPhe 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 TGGGCATTGCTGAGGGCTAGCACGGCTTTCAAGCTGTACTACCACAAGAGCCCCAAGTTC 3960

Qy       1321 ValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArgPro 1340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 GTGTGGCAGATGGCGGGCAGACAGCTCGCGTACATTAAGGCGCAGATGACGAGCAGACCC 4020

Qy       1341 GlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLys 1360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 GGTGAAGGCGCCCCGGCGTTGATGACCGCGTTCATGTATGCGGGCTTGATGCCGGATAAG 4080

Qy       1361 LysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAspPro 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 AAGTTTACGAAGCAGTATGTGGCCAGGCTGGAGGGCGATGGATCGGAGTACCCTGATCCG 4140

Qy       1381 GluValTyrGluValLeuGlyAspAspAspPheAspGlyIleGlyPheThrGlyAsnGly 1400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 GAGGTCTATGAAGTGCTGGGCGATGATGATTTTGATGGAATTGGTTTCACAGGGAATGGC 4200

Qy       1401 AspTyr 1402
              ||||||
Db       4201 GATTAT 4206
```

```
RESULT 2
AX034425
LOCUS       AX034425                8045 bp    DNA     linear   PAT 22-SEP-2000
DEFINITION  Sequence 1 from Patent WO0050581.
ACCESSION   AX034425
VERSION     AX034425.1  GI:10303130
KEYWORDS    .
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Cogoni,C. and Macino,G.
  TITLE     Isolation and characterization of a n. Crassa silencing gene and
            uses thereof
  JOURNAL   Patent: WO 0050581-A 1 31-AUG-2000;
            UNIV ROMA (IT) ; COGONI CARLO (IT) ; MACINO GIUSEPPE (IT)
FEATURES             Location/Qualifiers
     source          1..8045
                     /organism="Neurospora crassa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5141"
     CDS             2447..6655
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC10121.1"
                     /db_xref="GI:10303131"
                     /db_xref="REMTREMBL:CAC10121"
                     /translation="MNPITPRKRNSPVEEIINRLNNDYNLGLQCVADTTLTPHRRKEL
                     AESDEDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLP
                     WSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNT
                     KRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQC
                     TKRAKGKLSDNVAAAAAPPVPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQ
                     GTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKS
                     HIDNQPLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAP
                     VVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDV
                     TDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTA
                     PLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQ
                     WLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFRP
                     DVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKTYAI
                     MTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFG
                     SAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQL
                     LPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRV
                     SHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRV
                     GRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQ
                     RVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEM
                     PLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNY
                     KERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDP
                     MYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWDPDLA
                     SYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQVY
                     EKWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHKSPKF
                     VWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGDGSEYP
                     DPEVYEVLGDDDFDGIGFTGNGDY"
ORIGIN

Alignment Scores:
Pred. No.:              0              Length:          8045
Score:                  7397.00        Matches:         1402
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            100.00%        Indels:          0
DB:                     6              Gaps:            0

US-09-913-878A-2 (1-1402) x AX034425 (1-8045)

Qy          1 MetAsnProIleThrProArgLysArgAsnSerProValGluGluIleIleAsnArgLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 ATGAACCCTATTACTCCTAGGAAGAGGAATAGCCCCGTCGAGGAAATCATAAACCGGCTC 2506

Qy         21 AsnAsnAspTyrAsnLeuGlyLeuGlnCysValAlaAspThrThrLeuThrProHisArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 AATAACGACTACAACCTGGGCCTCCAGTGTGTCGCAGACACAACTCTCACCCCCCACCGC 2566

Qy         41 ArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHisAspLysIleTyrArgAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 CGGAAGGAGCTGGCCGAGAGTGACGAGGATTTCGGTCGCCATGACAAGATCTACAGAGCC 2626

Qy         61 LeuAsnPheLeuTyrTrpArgLysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2627 CTGAACTTTCTCTACTGGCGGAAGGATGACTCCCTGAACCAGGCAGAAGCCAACTTCTTC 2686

Qy         81 IleGluAlaLysAlaAlaSerSerAsnTrpValProLysAlaHisAlaAspProAspThr 100
Db       2687 ATCGAGGCCAAAGCTGCGAGCTCGAACTGGGTGCCCAAAGCCCACGCCGACCCTGACACG 2746

Qy        101 LeuProTrpSerLysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGln 120
Db       2747 CTTCCGTGGTCCAAGGAACCTCCCCGCGCCGCTACTGCCGGCCAACAATGGGCATTGCAG 2806

Qy        121 ThrValLeuLeuGluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArg 140
Db       2807 ACTGTGTTGCTCGAGGTGCTTAATAGGTTTATGCCACCTCCCAATAACACACCAGGTCGA 2866

Qy        141 ThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThr 160
Db       2867 ACGTTTGGCAGAACTCTAAGCGGCCCAAGTGGCCTGAGCCGCCCAACCTCTACCAACACC 2926

Qy        161 LysArgLysAspGluProAlaAsnValThrPheAlaAspProProLysArgSerLeuThr 180
Db       2927 AAACGCAAGGATGAGCCCGCCAATGTCACTTTCGCTGATCCGCCCAAACGCTCGTTGACT 2986

Qy        181 ArgSerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspPro 200
Db       2987 CGCTCTGCCACAGGTCCTCCTATTCACGGCGCGGCGATACCCCTAAAGTTCCCCGATCCA 3046

Qy        201 ValAsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLys 220
Db       3047 GTGAATACCGGTTCCAAACGACCATCTCTCGAGAGTGAGAATCTCAATCAGTGCACCAAG 3106

Qy        221 ArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValProIle 240
Db       3107 CGGGCCAAGGGCAAGCTGTCTGATAATGTTGCCGCTGCCGCCGCCCCGCCCGTGCCTATT 3166

Qy        241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
Db       3167 GCGAGCGCTTTGGACAAGGTACCGACTCGAAGGCATGCCAATACGAGAGATCCCACGGCG 3226

Qy        261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
Db       3227 ACAGGTCATAGACGAGCGGACCAGGTGGATTCCTTTGATACATCTCAAGGCACTTCCTAT 3286

Qy        281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
Db       3287 GGTTCGAGTGTCTTCAGCGCTTGCCGTCACAATCAGAGCACTACCCAGAGTAGTTTTGAG 3346

Qy        301 AlaProProSerGlnProArgGluLysArgProValAspAlaThrValPheGluAlaGly 320
Db       3347 GCTCCTCCTTCACAGCCCAGAGAGAAGCGGCCTGTGGATGCCACGGTCTTTGAGGCTGGA 3406

Qy        321 HisLeuIleGluSerProSerLysGlyArgThrThrLysSerHisIleAspAsnGlnPro 340
Db       3407 CACTTGATTGAGTCTCCTAGCAAAGGAAGAACAACCAAGTCCCACATAGATAACCAGCCC 3466

Qy        341 LeuSerSerSerSerGlnGlyGluThrSerPheSerThrTyrTyrGluSerPheProSer 360
Db       3467 CTTTCATCGTCTTCCCAGGGTGAAACTTCGTTCAGCACTTACTATGAGTCGTTTCCAAGT 3526

Qy        361 SerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLeuAlaArgSerGlu 380
Db       3527 TCCGGCGGCGAGGGCGCAATTCCCGAGCCGAGTCGCTCAAATGGACTGGCTCGGAGCGAA 3586

Qy        381 GluSerAlaArgSerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsn 400
Db       3587 GAAAGCGCTCGATCTCAGGTTCAAGTTCATGCCCCGGTGGTTGCAGCTCGGCTGAGAAAT 3646

Qy        401 IleTrpProLysPheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluVal 420
Db       3647 ATTTGGCCGAAATTTCCCAAATGGCTACACGAAGCTCCTCTCGCTGTTGCATGGGAAGTT 3706

Qy        421 ThrArgLeuPheMetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyr 440
Db       3707 ACCAGACTCTTTATGCACTGCAAAGTAGACTTGGAAGACGAGAGCCTGGGCCTAAAGTAC 3766

Qy        441 AspProSerTrpSerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeu 460
Db       3767 GACCCTTCCTGGTCTACCGCGCGCGATGTCACAGATATCTGGAAGACTCTCTACCGGCTT 3826

Qy        461 AspAlaPheArgGlyLysProPheProGluLysProProAsnAspValPheValThrAla 480
Db       3827 GATGCTTTCCGTGGTAAACCCTTTCCAGAAAAGCCGCCCAACGACGTGTTCGTGACCGCA 3886

Qy        481 MetThrGlyAsnPheGluSerLysGlySerAlaValValLeuSerAlaValLeuAspTyr 500
Db       3887 ATGACGGGCAACTTTGAGAGCAAAGGTAGTGCCGTTGTTCTCTCTGCTGTTCTAGACTAC 3946

Qy        501 AsnProAspAsnSerProThrAlaProLeuTyrLeuValLysLeuLysProLeuMetPhe 520
Db       3947 AATCCGGACAACTCGCCTACTGCGCCCCTTTACCTTGTGAAGCTGAAGCCGCTCATGTTC 4006

Qy        521 GluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIleLeuIle 540
Db       4007 GAGCAGGGCTGTCGACTCACCCGTCGGTTCGGTCCTGATAGGTTTTTCGAGATCCTTATA 4066

Qy        541 ProSerProThrSerThrSerProSerValProProValValSerLysGlnProGlyAla 560
Db       4067 CCGTCGCCTACGAGCACCAGCCCAAGTGTACCGCCGGTGGTCAGCAAACAACCAGGTGCG 4126

Qy        561 ValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrp 580
Db       4127 GTCGAAGAAGTCATCCAGTGGCTCACGATGGGGCAACATTCTCTTGTAGGCCGCCAATGG 4186

Qy        581 ArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArg 600
Db       4187 CGCGCTTTCTTCGCCAAAGATGCCGGATACAGGAAACCTCTCAGGGAGTTCCAGCTCCGC 4246

Qy        601 AlaGluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIle 620
Db       4247 GCCGAGGACCCGAAACCCATCATCAAGGAGAGAGTCCACTTCTTTGCCGAGACCGGCATT 4306

Qy        621 ThrPheArgProAspValPheLysThrArgSerValValProAlaGluGluProValGlu 640
Db       4307 ACGTTCCGACCTGATGTGTTCAAGACGAGATCTGTCGTTCCGGCAGAGGAACCTGTAGAG 4366

Qy        641 GlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAspAsnAsn 660
Db       4367 CAACGGACCGAGTTCAAAGTTAGTCAAATGCTGGACTGGCTCCTGCAACTCGACAACAAC 4426

Qy        661 ThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTyr 680
Db       4427 ACTTGGCAGCCGCACCTCAAGTTGTTCTCCCGTATCCAGCTCGGTCTGAGTAAGACATAT 4486

Qy        681 AlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAspLeuLeuSerPro 700
Db       4487 GCCATTATGACATTGGAGCCTCACCAGATCAGACACCACAAGACCGATCTTCTTTCACCT 4546

Qy        701 SerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSerValAlaLys 720
Db       4547 TCAGGCACTGGCGAAGTGATGAATGACGGTGTAGGCCGCATGTCGCGAAGCGTGGCCAAG 4606

Qy        721 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGly 740
Db       4607 AGGATACGCGATGTTCTCGGTTTGGGTGATGTGCCCTCTGCTGTGCAAGGGCGGTTTGGT 4666

Qy        741 SerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGlu 760
Db       4667 TCGGCCAAGGGAATGTGGGTTATCGACGTTGACGACACAGGCGATGAGGATTGGATCGAG 4726

Qy        761 ThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 780
Db       4727 ACATACCCGTCCCAGCGCAAGTGGGAATGCGACTTCGTTGATAAACATCAACGTACCCTC 4786

Qy        781 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuPro 800
Db       4787 GAAGTCCGGAGCGTGGCTTCTGAACTGAAGTCAGCTGGTCTCAACCTACAGCTGTTACCT 4846
```

```
Qy        801 ValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeu 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4847 GTCCTGGAAGATAGAGCCAGGGACAAGGTGAAGATGCGCCAGGCAATCGGTGACCGTCTT 4906

Qy        821 IleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4907 ATCAACGATTTGCAACGACAGTTCAGCGAGCAAAAGCATGCTTTGAATCGCCCAGTGGAA 4966

Qy        841 PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArg 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4967 TTTCGCCAATGGGTTTACGAGAGTTATTCCAGTCGCGCAACTCGAGTCAGCCACGGCCGT 5026

Qy        861 ValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsn 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5027 GTGCCTTTTCTTGCTGGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAAC 5086

Qy        881 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5087 AGTGGGTTCGATCCCAAGAAGCAAAAGTACTTGCAAGACATCGCCTGGGATCTTCAAAAG 5146

Qy        901 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5147 CGGAAATGTGACACGTTGAAGTCCAAGCTGAACATCCGTGTCGGTCGATCAGCATACATT 5206

Qy        921 TyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer 940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5207 TACATGATTGCCGATTTCTGGGGTGTGCTTGAGGAAAATGAGGTTCATGTCGGATTCTCC 5266

Qy        941 SerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5267 TCAAAGTTCAGAGACGAGGAGGAGTCTTTTACACTCCTATCGGACTGTGATGTCCTCGTG 5326

Qy        961 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5327 GCGCGATCCCCAGCCCATTTCCCTAGTGATATCCAACGGGTTCGAGCAGTCTTCAAGCCA 5386

Qy        981 GluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAla 1000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5387 GAGCTCCACAGTCTCAAGGATGTAATCATCTTCTCTACTAAAGGAGATGTACCGCTTGCT 5446

Qy       1001 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5447 AAGAAGCTATCTGGTGGAGACTACGACGGCGATATGGCCTGGGTCTGCTGGGATCCGGAG 5506

Qy       1021 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu 1040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5507 ATCGTCGATGGTTTCGTCAATGCGGAAATGCCTCTGGAGCCCGACCTGTCTAGGTACCTA 5566

Qy       1041 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 1060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5567 AAGAAGGACAAAACGACTTTCAAACAACTTATGGCCTCACACGGCACGGGCTCAGCGGCC 5626

Qy       1061 LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5627 AAAGAGCAGACTACATACGATATGATCCAGAAGAGCTTCCATTTCGCCCTGCAGCCCAAC 5686

Qy       1081 PheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSer 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5687 TTCTTGGGCATGTGCACTAACTACAAAGAAAGGCTCTGTTACATCAACAATAGTGTGTCT 5746

Qy       1101 AsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGln 1120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5747 AACAAGCCGGCCATCATTCTTAGTTCACTGGTGGGAAACCTCGTCGATCAGAGCAAGCAA 5806

Qy       1121 GlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAla 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5807 GGTATTGTCTTTAACGAAGCAAGCTGGGCTCAATTGCGTAGGGAACTGCTTGGCGGTGCA 5866

Qy       1141 LeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThr 1160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5867 TTGTCCCTTCCTGACCCAATGTACAAGAGCGACAGTTGGCTCGGGCGCGGAGAGCCTACC 5926

Qy       1161 HisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGlu 1180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5927 CACATTATTGACTACCTGAAATTCTCCATCGCCAGGCCTGCGATTGACAAGGAACTGGAA 5986

Qy       1181 AlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAsp 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5987 GCCTTCCACAATGCCATGAAAGCGGCCAAGGATACAGAAGACGGCGCTCACTTTTGGGAT 6046

Qy       1201 ProAspLeuAlaSerTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSer 1220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6047 CCGGATCTCGCTTCCTACTACACGTTCTTCAAGGAGATTAGCGACAAGTCGCGATCGTCC 6106

Qy       1221 AlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArg 1240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6107 GCACTGCTATTCACGACTCTGAAGAACCGTATCGGCGAAGTCGAGAAAGAATATGGCAGG 6166

Qy       1241 LeuValLysAsnLysGluMetArgAspSerLysAspProTyrProValArgValAsnGln 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6167 TTGGTCAAAAACAAGGAGATGAGAGACAGCAAGGACCCCTACCCTGTCCGCGTCAACCAG 6226

Qy       1261 ValTyrGluLysTrpCysAlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyr 1280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6227 GTTTATGAAAAATGGTGCGCCATCACGCCTGAGGCGATGGACAAATCCGGAGCAAATTAT 6286

Qy       1281 AspSerLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThr 1300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6287 GATTCTAAGGTGATCAGGTTGCTGGAGCTGTCCTTCCTCGCGGACCGTGAGATGAATACA 6346

Qy       1301 TrpAlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPhe 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6347 TGGGCATTGCTGAGGGCTAGCACGGCTTTCAAGCTGTACTACCACAAGAGCCCCAAGTTC 6406

Qy       1321 ValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArgPro 1340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6407 GTGTGGCAGATGGCGGGCAGACAGCTCGCGTACATTAAGGCGCAGATGACGAGCAGACCC 6466

Qy       1341 GlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLys 1360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6467 GGTGAAGGCGCCCCGGCGTTGATGACCGCGTTCATGTATGCGGGCTTGATGCCGGATAAG 6526

Qy       1361 LysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAspPro 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6527 AAGTTTACGAAGCAGTATGTGGCCAGGCTGGAGGGCGATGGATCGGAGTACCCTGATCCG 6586

Qy       1381 GluValTyrGluValLeuGlyAspAspAspPheAspGlyIleGlyPheThrGlyAsnGly 1400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6587 GAGGTCTATGAAGTGCTGGGCGATGATGATTTTGATGGAATTGGTTTCACAGGGAATGGC 6646

Qy       1401 AspTyr 1402
              ||||||
Db       6647 GATTAT 6652
```

RESULT 3
AP004357/c
LOCUS   AP004357   165701 bp   DNA   linear   PLN 27-NOV-2003
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC clone:B1074C08.
ACCESSION   AP004357 BA000010
VERSION   AP004357.4 GI:21328130
KEYWORDS   .
SOURCE   Oryza sativa (japonica cultivar-group)
  ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T., Yamamoto,K., Sakata,K., Baba,T., Katayose,Y., Wu,J., Niimura,Y., Cheng,Z., Nagamura,Y., Antonio,B.A., Kanamori,H., Hosokawa,S., Masukawa,M., Arikawa,K., Chiden,Y., Hayashi,M., Okamoto,M., Ando,T., Aoki,H., Arita,K., Hamada,M., Harada,C., Hijishita,S., Honda,M., Ichikawa,Y., Idonuma,A., Iijima,M., Ikeda,M., Ikeno,M., Itoh,S., Itoh,T., Itoh,Y., Itoh,Y., Iwabuchi,A., Kamiya,K., Karasawa,W., Katagiri,S., Kikuta,A., Kobayashi,N., Kono,I., Machita,K., Maehara,T.,

```
            Mizuno,H., Mizubayashi,T., Mukai,Y., Nagasaki,H., Nakashima,M.,
            Nakama,Y., Nakamichi,Y., Nakamura,M., Namiki,N., Negishi,M.,
            Ohta,I., Ono,N., Saji,S., Sakai,K., Shibata,M., Shimokawa,T.,
            Shomura,A., Song,J., Takazaki,Y., Terasawa,K., Tsuji,K., Waki,K.,
            Yamagata,H., Yamane,H., Yoshiki,S., Yoshihara,R., Yukawa,K.,
            Zhong,H., Iwama,H., Endo,T., Ito,H., Hahn,J.H., Kim,H.I., Eun,M.Y.,
            Yano,M., Jiang,J. and Gojobori,T.
  TITLE     The genome sequence and structure of rice chromosome 1
  JOURNAL   Nature 420 (6913), 312-316 (2002)
  MEDLINE   22337376
   PUBMED   12447438
REFERENCE   2  (bases 1 to 165701)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Jun 6, 2002 this sequence version replaced gi:20975426.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'.  A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from M13rev to -21M13 of the BAC
            clone. This sequence of B1074C08 clone has an overlap with B1147B04
            clone (DDBJ:AP004641) at the position 1 to 108,335 of 5' end and an
            overlap with B1129H01 clone (DDBJ: AP003370) at the position
            115,348 to 165,701 of 3' end. Detailed information on overlap and
            assembly quality together with annotation of this entry is
            available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1. .165701
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="1"
                     /clone="B1074C08"
     gene            complement(join(1956. .2373,2716. .2783))
                     /gene="B1074C08.1"
     CDS             complement(join(1956. .2373,2716. .2783))
                     /gene="B1074C08.1"
                     /note="contains EST D22801(C1352)
                     unknown protein"
                     /codon_start=1
                     /protein_id="BAC00710.1"
                     /db_xref="GI:21328131"
                     /translation="MDSTLTSQKATATATFIITSTVRINRFLGRTSGATKKPKGASFQ
                     AMESYVLTWFIVFAFDSSDMVIVQCTHSRTLELKSYFCYCYCYLYHFFLLASMYKRPT
                     PFWHSTQPKKNQNFPRRSDPTNEATNRPTEEDRGMDLEDRVGGDQSCQRRSMTTQRWM
                     M"
     gene            join(7311. .7428,7640. .7776)
                     /gene="B1074C08.2"
     CDS             join(7311. .7428,7640. .7776)
                     /gene="B1074C08.2"
                     /note="contains EST AU069673(C0828)
                     unknown protein"
                     /codon_start=1
                     /protein_id="BAC00711.1"
                     /db_xref="GI:21328132"
                     /translation="MHCHCVRCNVRSGYAQQQPFVRACVRAAQPRWKVHSNTRPAARM
                     KLIGSGSGSGGQRRGIARALKEHKARLYIIRRCVVMLLRC"
     gene            join(9087. .9169,9577. .9836,9998. .10158,10204. .10298,
                     10342. .10570)
                     /gene="B1074C08.3"
     CDS             join(9087. .9169,9577. .9836,9998. .10158,10204. .10298,
                     10342. .10570)
                     /gene="B1074C08.3"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAC00712.1"
                     /db_xref="GI:21328133"
                     /translation="MARQQCLDRVGRDGDGAANAVEKMREREVTDMWGPPVGPFFHLP
                     PPLPKPATTAPTPLAYPAPPRLHSRTPAPSPRPRRLPRPSLPAHARAPKRAGFDLNPP
                     PTLSLSLSIPHVGPPPSRAVIAFGFACSRAISSTSPLSKSAIGGRHSPCAGEPFHRRR
                     RRPWFSLTVVLELVRSLPPFPGRSPSLFAGVFPHAGRIPRAEPRRRPSRHGLRPTLPP
                     PAPPVTTRAAAASHGRLVGFEAEPSPAAAPRAVRSARGRSGQSSCGPAVDHPRGPGP"
     gene            join(16765. .17574,19784. .20080)
                     /gene="B1074C08.4"
     CDS             join(16765. .17574,19784. .20080)
                     /gene="B1074C08.4"
                     /codon_start=1
                     /product="putative DNA damage repair protein"
                     /protein_id="BAC00713.1"
                     /db_xref="GI:21328134"
                     /translation="MLGGLYGDLPPPSSSADDDKPSAAGWSSAAKMAPPTLRKPPATF
                     APPPSVLRNSRPAPKAPAAQPPPPPTLPIETTTSTSFQPALVAVQSTVMEEYDPARPN
                     DYEDYRKDKLKRAKEAEVRKELERRRREEEERERERELREREGRDALNISGEEAWKRR
                     AAMSGSAAPRPSSSPPHGDGFAIGNSSSSGLGLGAGGQMTAAQRMMARMGWKEGQGLG
                     KQEQGITAPLVAKKTDRRGGVIVDENSSKQEKKPKSVNFDGPPTRVLLLRNMVGPGEV
                     DDELEEEVASECSKYGTVLRVLIFEITQADFPAEEAVRIFILFERAEEATKAMIDLEG
                     RFFGGRVVRATFFDEERFGKNQLAPMPGEVAGFD"
     gene            join(22181. .22363,22471. .22640,22731. .22803,22919. .22993,
                     23119. .23229,23337. .23382,23475. .23538,24200. .24302,
                     24389. .24484,24576. .24605,24824. .24923,25151. .25235,
                     25355. .25478)
                     /gene="B1074C08.5"
     CDS             join(22181. .22363,22471. .22640,22731. .22803,22919. .22993,
                     23119. .23229,23337. .23382,23475. .23538,24200. .24302,
                     24389. .24484,24576. .24605,24824. .24923,25151. .25235,
                     25355. .25478)
                     /gene="B1074C08.5"
                     /note="contains ESTs
                     C26310(C12073),D15391(C0569),AU100694(C12073)
                     unknown protein"
                     /codon_start=1
                     /protein_id="BAC00714.1"
                     /db_xref="GI:21328135"
                     /translation="MAPGTTLAPKRRKAEASPSPSPSPMGDSSDGGYSDSDLHDAEES
                     FYSARSGSEDDRQVSSSNDDDDSEEEEQEEREMDEEEDEEDDDDEEMNEEDDEDEGEE
                     MNELEKEYRTLQTNQQNILETLKQHRDDDVSKGQAVKNQKVLWDKALEMRFLLQKAFS
                     TSNKLPKEPIRSMFCDHNQEIEQAYLDLLNSSKQTLGSMMELQEALLERNRATKDVTD
                     TDNSSELNGEDDEWSEVQKLQKRITPFRNSEIDKWQRKTQVTTGAAALKGKLHAFNQN
                     ISDQVTSYMRDPSRMINRMHLRKSTLGVFGEEVGEHENNKEENNTEGDPELVDDSEFY
                     QQLLKEFLESCDAGASESAFYALKKQQHKKRKLVDRRASKSRKIRYHVHEKIANFMAP
                     VPMVIPPMAPKLFENLFGMGNQKSTTA"
     gene            join(26433. .26525,30039. .30091,31996. .32168,32606. .32724)
                     /gene="B1074C08.6"
     CDS             join(26433. .26525,30039. .30091,31996. .32168,32606. .32724)
                     /gene="B1074C08.6"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAC00715.1"
                     /db_xref="GI:21328136"
                     /translation="MFIHRTKLLDAAAQLAVISDRRVFVYTCLMKPSSSSFVVDTISP
                     LSQHRGVTPLLLAVRSRGLAHFFFHRDQPGSLLTPAKSAVMWLTPDVQQATHIAHLTQ
                     LRSVQSTTTEKLIGAGWEPALGAGFPNRHHGGDTDDFQSRCRF"
     gene            36236. .36550
                     /gene="B1074C08.7"
     CDS             36236. .36550
                     /gene="B1074C08.7"
                     /note="hypothetical protein"
                     /codon_start=1
```

```
                     /protein_id="BAC00716.1"
                     /db_xref="GI:21328137"
                     /translation="MAAALGQIWRRRGAAAAFPPATTRRRCRRLPSRYHPPPPPPPL
                     LPGCRYRSYLVEGRRCRRRHRLPSCHHPPSPSPPPLLPGRHSRPDLVEDAASTAGPPP
                     RV"
     gene            complement(38211..39233)
                     /gene="B1074C08.8"
     CDS             complement(38211..39233)
                     /gene="B1074C08.8"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAC00717.1"
                     /db_xref="GI:21328138"
                     /translation="MDWASLPADLLFSISSHLREPEDFVRFRAVCPQWRAAVSHKEHA
                     FFQPWIMASRWLEDEYSENVLFYSLSTLKTIKVHVPDMKSRRIAASGSSHLIAIDKDD
                     DLSAVLINPLSGKTTALPRLPKFFHDNGAHGWITGEGVITVVLNNWMSENMALWYRGG
                     GITMKGWAVVPRWKLRLRVSHYLRKLAAYGDQMEMHLIDLGGDNEDSVVLLQETKKVE
                     LLGGCWPRSDEVFKATRPCHHEWFSLYRNVEQEEIPVHDIGNVMVVQSRDSCTRTYMI
                     PASRDFAALGSRNAFYYLWKQFDAGGSYNALFKKCLASEVLTFVKRLPEDWKLSDEWF
                     MPSLKY"
     LTR             42943..43587
                     /note="5' LTR"
     gene            join(44015..44093,44667..44836)
                     /gene="B1074C08.9"
     CDS             join(44015..44093,44667..44836)
                     /gene="B1074C08.9"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAC00718.1"
                     /db_xref="GI:21328139"
                     /translation="MPPRRCLEAAAPRRVPRRASLPCAAAEPEGVADLSEGEDPQDYE
                     QGKSSLFFEHVDPNCEIILFTNKIACNDEHPTCDYVMP"
     gene            join(47189..47443,47555..47728)
                     /gene="B1074C08.10"

Alignment Scores:
Pred. No.:               0.000141       Length:         165701
Score:                   348.50         Matches:        242
Percent Similarity:      35.55%         Conservative:   154
Best Local Similarity:   21.72%         Mismatches:     383
Query Match:             4.71%          Indels:         337
DB:                      8              Gaps:           51

US-09-913-878A-2 (1-1402) x AP004357 (1-165701)

Qy        102 ProTrpSerLysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThr 121
               |||     :::  |||  |||        ||||||||            |||
Db     113663 CCTGAGGGGGAAGCCCCTGCGCGTCGCGTCCGCGCCGGACAG---------CTCGCTCCG 113613

Qy        122 ValLeuLeuGluValLeuAsnArgPheMetProProProAsnAsnThrPro--------- 138
                 :::             :::|||  :::|||  |||        |||
Db     113612 CGTGTCCCGCCGCAGCAGCATCGCGCCGTTCCGGTTCCCCGACGTGCGCCTCGAGGTCGG 113553

Qy        139 GlyArgThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArg---------Pro 155
               |||      |||      |||  |||  |||      |||
Db     113552 GGCGCTCCCGTCGCCCGGCGCGT-TCCTCGCCGCCTGGCGCGGGCCGGACGCCGGGCTCG 113494

Qy        156 ThrSerThrAsnThrLysArgLysAspGluProAlaAsnValThrPheAlaAspProPro 175
              ||||||    :::|||              |||:::  :::   |||   ||||||
Db     113493 ACCTCTCCGTCGACCCCTTCGACGGGTGCTGCCGCCTCGTGTTCACCCGCGACACCGCCT 113434

Qy        176 LysArgSerLeuThrArgSerAlaThrGlyProProIleHisGlyAlaAlaIleProLeu 195
                 ||||||         ||||||    ||||||              |||
Db     113433 TCACGTTCC---CGGGCTTCCGCGAGGTCGCCGCCATAC---------GCTGCGACGTCA 113386

Qy        196 LysPheProAspProValAsnThrGlySerLysArgProSerLeuGluSerGluAsnLeu 215
                 :::      |||      |||      :::
Db     113385 AGCTGGAGTTCCCCGTGCGCGACGTCCTGGAGG--------------------------- 113353

Qy        216 AsnGlnCysThrLysArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAla 235
               :::  ||||||                  |||            ::::::|||:::
Db     113352 TCCGCCTGTACC------------------GCCTCGACTGCTCGCTGCTCCTCCGCCTCG 113311

Qy        236 ProProValProIleAlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThr 255
              ||||||                                    ||||||         |||
Db     113310 CCGCCG------------------------------------CGCCGCTGGTGCACTACC 113287

Qy        256 ArgAspProThrAlaThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSer 275
                 ||||||   |||         |||:::                |||      |||:::
Db     113286 GCACCGCCGACGACGACTTCCACGAGCCGGTCCCCTTCGACCTGCTCGACGACGACGACC 113227

Qy        276 GlnGlyThrSerTyrGlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThr 295
              :::|||               :::|||      :::|||                  |||
Db     113226 CGTGGATACGGACCACGGACATCACCCCGAGCGGCGCCATCGGGCGGTGCGGGGTGTACA 113167

Qy        296 GlnSerSerPheGluAlaProProSerGlnProArgGluLysArgProValAspAlaThr 315
                 ||||||   |||               |||:::      |||   :::|||
Db     113166 GG-ATCTCCTTCTCGGCG---CGATTCTGGCCCAAGATGGACCGCGCGCTGGACTACATG 113111

Qy        316 ValPheGluAlaGlyHisLeuIleGlu---------------SerProSerLysGlyArg 330
                             :::::::::                  |||   :::|||
Db     113110 AGGGAGAGGAGGGTGGCGATCGTCGATTGCGGCGGCGGATGGGGGCCGCGGCGGGGGCTC 113051

Qy        331 ThrThrLysSerHisIleAsp---AsnGlnProLeuSerSerSerSerGlnGlyGluThr 349
              |||   :::      ::::::      :::|||:::                  :::
Db     113050 ACAGTGCGCGACGAGCTGGAGTTCGGGGAGCCAATGCAG--------------GACGTC 113006

Qy        350 SerPheSerThrTyrTyrGlu---------SerPhePro------------------Ser 360
                 |||      :::               ||||||                   :::
Db     113005 TTCTTCTGCCTGCAGCACGCCGAGGGGCTCAAGTTCCCGCTCCTCTTCATGGTGAACGCG 112946

Qy        361 SerGlyGlyGluGlyAlaIleProGlu------------ProSerArgSerAsnGlyLeu 376
                       :::|||   |||   :::            |||         :::   |||
Db     112945 TTGGTGCACAAGGGAATCATCAACCAGCACCAGCTCACGCCTGAATTCTTCAGCCTGCTG 112886

Qy        377 AlaArgSerGluGluSerAlaArgSerGlnValGlnValHisAlaProValValAlaAla 396
                 ||||||||||||:::                               |||   |||
Db     112885 GGGAGAAGCGAGGAGAAC-----------------------------GTCAATGTGGCT 112856

Qy        397 ArgLeuArgAsnIleTrp---ProLysPheProLysTrp--------------LeuHis 410
                 ||||||:::   |||   |||||||||   :::                  |||
Db     112855 GCATTGAGGGATTTCTGGGGGGACAAGTTCCCAGTGTTTGATGCGTGCGGGAGGCTGAAG 112796

Qy        411 GluAlaProLeuAlaValAla---------------------------------------- 417
              :::|||      ||||||
Db     112795 AAGGCACTGAATCGGGTGGCCAGGAACCCCAAACTTCTCTGCAGCAAGGTCGGGGATGAC 112736

Qy        418 ---TrpGluValThrArgLeuPheMet---------------HisCys---LysValAsp 430
                 ||||||   ||||||   :::               :::|||
Db     112735 CACGCGGAGGTGCGGAGGCTGGTGATCACGCCCACCAGGGCTTATTGTCTGCCTCCAGAA 112676

Qy        431 LeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrpSerThrAlaArgAspVal 450
              :::|||      :::      |||::::::
Db     112675 GTGGAGCGCTCAAACCGTGTTCTTCGGCAT------------------------------ 112646

Qy        451 ThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArgGlyLysProPheProGlu 470
                                                                 :::   |||
Db     112645 -------------------------------------------------TACCATGAG 112637

Qy        471 LysProProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSerLysGlySer 490
                          :::   ||||||   |||
Db     112636 GTGGCTGACAGGTTTTTGAGGGTCACTTTTATGGACGAGGGTATGCAG------------ 112589

Qy        491 AlaValValLeuSerAlaValLeuAspTyrAsnProAspAsnSerProThrAlaProLeu 510
                                ||||||:::   |||      ||||||   |||||||||:::
Db     112588 -----------------GTGCTGAACAACAATGTGCTCAACTCCTTCACTGCACCAATT 112547

Qy        511 TyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrArgArgPhe 530
                          :::|||   ||||||
Db     112546 ------------GTCAAAGACTTGATG--------------------------------- 112532

Qy        531 GlyProAspArgPhePheGluIleLeuIleProSerProThrSerThrSerProSerVal 550
                          ||||||:::
Db     112531 ------TCGAATTTTTTCCAG--------------------------------------- 112517

Qy        551 ProProValValSerLysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMet 570
```

```
Db   112516 ---------------CAGAAGACAACGGTGTACAAGCGTGTCAGGATGTTGTTGACGGAG 112472
Qy      571 GlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyr 590
Db   112471 GGTTTCCAC---ATGTGTGGAAGGAAGTAC---TCATTTCTCGCATTCTCATCGAAC--- 112421
Qy      591 ArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysProIleIleLysGlu 610
Db   112420 ---------------------CAGTTAAGG------------------------GACAAG 112406
Qy      611 ArgValHisPhePheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArg 630
Db   112405 TCAGCTTGGTTCTTTGCCGAG--------------------------------------- 112385
Qy      631 SerValValProAlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMet 650
Db   112384 ---------------------------GACAGAAAGACA------ACGGTGGAAGCAATT 112358
Qy      651 LeuAspTrpLeuLeuGlnLeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSer 670
Db   112357 AGGAAGTGGATGGGACGGTTCACAAGTAAGAATGTTGCGAAGCATGCT---------GCT 112307
Qy      671 ArgIleGlnLeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIle 690
Db   112306 CGAATGGGGCAGTGCTTCTCATCCACATATGCAACTGTGACAATGCGGCCGGATGAGGTT 112247
Qy      691 ArgHisHisLysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGly 710
Db   112246 GATGAGAGTTTTGATGATGTTGTGCAT---------AATGAGTACATTTTCTCCGATGGA 112196
Qy      711 ValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAsp 730
Db   112195 ATTGGCAAAATTACCCCAGATCTTGCATTGGAAGTTGCCGAGAGGCTGCAGCTGACAGAT 112136
Qy      731 ---ValProSerAlaValGlnGlyArgPheGlySerAlaLysGly---------MetTrp 746
Db   112135 AACCCGCCATCTGCTTATCAGATCAGGTTTGCTGGCTTCAAGGGTGTCATAGCTGTCTGG 112076
Qy      747 ValIleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArg 766
Db   112075 ------------CAAGGACATGGTGATGGGACACGGCTTTTCCTGAGGCCAAGCATGAGG 112028
Qy      767 LysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAla 786
Db   112027 AAGTTTGAGTCTAACCATTTG---------------GTGTTAGAGGTGGTCTCC---TGG 111986
Qy      787 SerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAla 806
Db   111985 ACAAAGTTCCAGCCAGGATTCTTAAATCGACAGATTATAATATTGCTATCCTCA------ 111932
Qy      807 ArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArg 826
Db   111931 --------------------CTGAATGTCCCAGATTCTATCTTTTGGCAAATGCAAGAG 111893
Qy      827 GlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyr 846
Db   111892 ACCATGCTTTCTAAC---------CTCAACAATATTCTATCAGACAGAGATGTTGCTTTT 111842
Qy      847 GluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGly 866
Db   111841 GAGGTTTTAACAACTTCATGTGCT------------------------------------ 111818
Qy      867 LeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLys 886
Db   111817 ---------GATGATGGAAACACTGCAGCATTGATGCTCAGTGCTGGCTTTGAACCTAGA 111767
Qy      887 LysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeu 906
Db   111766 ACTGAACCACACTTGAAAGCAATGCTCTTGGCTATAAGGTCTGCACAATTGCAGGATCTT 111707
Qy      907 LysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPhe 926
Db   111706 TTGGAAAAAGCAAGGATATTTGTGCCAAAGGGAAGGTGGTTGATGGGCTGTCTTGATGAG 111647
Qy      927 TrpGlyValLeuGluGlu------------------------------------AsnGlu 934
Db   111646 CTTGGGGTTCTTGAGCAAGGGCAGTGCTTTATTCGGGCAACAGTTCCATCATTGAATAGT 111587
Qy      935 ValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeuSer 954
Db   111586 TATTTTGTTAAGCATGGGTCAAGATTTTCATCAACAGATAAAAACACAGAGGTCATTTTG 111527
Qy      955 AspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgVal 974
Db   111526 GGT---ACTGTGGTAATAGCAAAGAATCCCTGTCTTCATCCAGGGGATGTCCGCATCCTT 111470
Qy      975 ArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLys 994
Db   111469 GAAGCAGTTGATGTGCCCGAACTGCATCATCTGGTTGATTGTTTGGTGTTCCCCCAGAAA 111410
Qy      995 GlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrp 1014
Db   111409 GGTGAGAGGCCACATGCTAACGAGGCATCTGGGAGCGATCTTGATGGGGATCTCTACTTT 111350
Qy     1015 ValCysTrpAspProGluIleValAsp------------------------------- 1023
Db   111349 GTGACATGGGATGAGAAACTTATACCTCCAGGCAAGAAGAGCTGGAACCCTATGGACTAC 111290
Qy     1024 ------------------------------------------------GlyPheValAsn 1027
Db   111289 TCCCCACCTGAAGCAAAACAACTTCCGCGCCAAGTATCTCAACATGTAAGTTCAATAACT 111230
Qy     1028 AlaGluMetPro------------------LeuGluProAspLeuSerArgTyrLeuLys 1041
Db   111229 CTTGAGATGTCAGCAATGTCCCATAATTTTTGTTTCCTACATACATGTCCTTTTTATAC 111170
Qy     1042 LysAspLysThrThrPhe---------------------------------------- 1047
Db   111169 GATGGATAACCCTCCTTTCATTGGAAGCATATATGGGTTTTCTTTACAGAACTGAGATAA 111110
Qy     1048 ---------LysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThr 1064
Db   111109 AGAAATTAAACGCAAGCTATGGTCGGGCAC---ACTGGCACACCAGCAAAAACAGGGAAT 111053
Qy     1065 ThrTyr---------AspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPhe 1081
Db   111052 AGCTTTAGAATCTCAGATATATTACAGCAATCA---------AAATTAAGTCCAAACCAG 111002
Qy     1082 LeuGly---------MetCysThrAsnTyrLysGluArgLeu 1092
Db   111001 AGAGGTCCTCCTAGCATTTGTAAACCTTATAGGCATCATCTC 110960
```

RESULT 4
NTA011576

```
LOCUS       NTA011576               3505 bp    mRNA    linear   PLN 25-FEB-2003
DEFINITION  Nicotiana tabacum mRNA for RNA-directed RNA polymerase.
ACCESSION   AJ011576
VERSION     AJ011576.1  GI:4138281
KEYWORDS    dRP gene; RNA-directed RNA polymerase.
SOURCE      Nicotiana tabacum (common tobacco)
  ORGANISM  Nicotiana tabacum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Nicotiana.
REFERENCE   1
  AUTHORS   Schiebel,W., Pelissier,T., Riedel,L., Thalmeir,S., Schiebel,R.,
            Kempe,D., Lottspeich,F., Saenger,H.L. and Wassenegger,M.
  TITLE     Isolation of an RNA-directed RNA polymerase-specific cDNA clone
            from Tomato
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3505)
  AUTHORS   Wassenegger,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1998) Wassenegger M., Max-Planck-Institut of
            Biochemistry Viroidresearch, Am Klopferspitz 18A, 82152
```

```
               Martinsried, GERMANY
FEATURES             Location/Qualifiers
     source          1..3505
                     /organism="Nicotiana tabacum"
                     /mol_type="mRNA"
                     /cultivar="Petit Havana SRI"
                     /db_xref="taxon:4097"
                     /tissue_type="leaves"
     gene            1..3505
                     /gene="RdRP"
     CDS             20..3370
                     /gene="RdRP"
                     /codon_start=1
                     /evidence=experimental
                     /product="RNA-directed RNA polymerase"
                     /protein_id="CAA09697.1"
                     /db_xref="GI:4138282"
                     /db_xref="GOA:Q9ZRY7"
                     /db_xref="SPTREMBL:Q9ZRY7"
                     /translation="MGKTIQVFGFPYLLSAEAVKKFLENHTGNGTVYALEVKQSKGGR
                     RAFAKVQFANNKSAEFIIDLASKGLYYGPSYLKAWEMKTDIVQPRTYVEHMDGVTLNF
                     GCQISDKKFAVLGSKDVSIKFGIGLKKIYFFLSHASADYKLQLSYENIWQVVLHRPYG
                     QNAQFLLLQLFGAPRIYKRLEDSCYSFFKETPDDQWVRTTDFTPSWIGLSSSLCLEFR
                     NGVQLPNFSESFFYYKESMNQFILQTGFTFSFSQKLALVPIVHLPEGIELPYKILFKI
                     SSLIQHGCFPGLALNFNFFQLVDPRRRNIACIEHALEKLYYLKECCYDPVRWLTEQYD
                     EYLKGRQLPKSPPITLDDGLVYVRRVVVTPCKVYFCGPEVNVSNRVLRNYSEDINNFL
                     RVSFVDEEWEKIHSTDLLPRASTGNGTRTDIYERILSTLRNGFIIGDKRFEFLAFSSS
                     QLRDNSVWMFASRPGLTANDIRTWMGDFRQIRNVAKYAARLGQSFGSSRETLSVGRHE
                     VEVIPDVACSLHGTNYIFSDGIGKISADFAHRVTIKCGLQYTPSSFQIRYGGYKGVVA
                     VDPYSSMKLSLRKSMLKYESNNIKLDVLGWSKYQPCYLNRQLVTLLSTLGVIDDVLEQ
                     KQNEAVDQLDAILHDSLKAQEALELMSPGENTNILKEMLNCGYMPDAEPFLSMMLQTF
                     RASKLLDLRTRTRIFIPNGRTMMGCLDESRTLEYGQVFVQFSGAGRRQFFEESHPFND
                     SGSANCDFILKGNVVVAKNPCLHPGDIRVLRAVDVPALHHMVDCVVFPQKGKRPHPNE
                     CSGSDLDGDIYFVCWDPDLIPPRQVQSMDYTPAPTTQLDHDVTIEEVEEYFTNYIIND
                     SLGIIANAHVVFADREPDMAMSDPCKQLAQLFSIAVDFPKTGVPAEIPSQLRPKEYPD
                     FMEKPDKPTYPSEKVIGKLFQKVKNKTPQASSIATFTRDVARKSYDSDMIVDGFEDYI
                     DEAFYYKSEYDNKLGNLMDYYGIKTEAEILSGGIMKASKTFDRRKDAEAIGVAVRCLR
                     KEARAWFKRRSDIDDMLAKASAWYHVTYHHTYWGLYNEGLKRDHFISFPWCVYDQLIQ
                     IKKAKARKRPVPHLSSLGSQLSRKLVIW"
ORIGIN

Alignment Scores:
Pred. No.:              7.77e-06        Length:         3505
Score:                  328.00          Matches:        210
Percent Similarity:     34.06%          Conservative:   151
Best Local Similarity:  19.81%          Mismatches:     385
Query Match:            4.43%           Indels:         314
DB:                     8               Gaps:           42

US-09-913-878A-2 (1-1402) x NTA011576 (1-3505)

Qy      441 AspProSerTrpSerThrAlaArgAspValThrAspIleTrpLysThrLeu--------- 457
              |||      |||            |||   |||       |||     |||
Db      596 GATGATCAGTGGGTAAGGACAACAGATTTCACTCCATCTTGGATAGGGCTATCTTCTAGC 655

Qy      458 -----------------------TyrArgLeuAspAlaPheArgGlyLysProPhePro 469
                                   :::|||         |||             |||
Db      656 TTATGTTTGGAGTTTCGTAATGGTGTTCAACTTCCAAATTTCAGCGAAAGTTTTTTTTAC 715

Qy      470 GluLysProProAsnAspValPheVal------ThrAlaMetThrGlyAsnPheGluSer 487
               |||             :::  |||:::         |||    |||     :::|||
Db      716 TATAAAGAAAGTATGAACCAATTTATTTTACAGACTGGTTTCACCTTTTCATTCTCTCAA 775

Qy      488 LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsnProAspAsnSerProThr 507
            |||   :::   |||   :::                 :::   |||:::
Db      776 AAATTGGCTCTGGTTCCCATTGTC------------CACCTTCCTGAGGGAATTGAGTTG 823

Qy      508 AlaProLeuTyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCys------Arg 525
                     |||   |||:::      |||:::   :::      |||||
Db      824 CCCTACAAGATATTGTTCAAAATTAGTTCCTTGATA---CAACATGGATGCTTTCCTGGG 880

Qy      526 LeuThrArgArgPheGlyProAspArgPhePheGluIleLeuIleProSerProThrSer 545
             |||         |||               |||||:::::::::  |||        :::
Db      881 CTAGCTTTAAATTTT---------AACTTTTTCCAATTAGTTGACCCTCGAAGGAGAAAT 931

Qy      546 ThrSerProSerValProProValValSerLysGlnProGlyAlaValGluGluVal--- 564
                                    ::::::               |||:::|||::::::
Db      932 -----------------------ATTGCATGCATTGAACACGCCTTGGAGAAACTGTAC 967

Qy      565 --------------------------IleGlnTrpLeuThr---MetGlyGlnHisSer 574
                                      ::::::|||||||||
Db      968 TATTTAAAGGAGTGTTGTTATGATCCAGTGAGGTGGCTAACTGAGCAGTATGATGAGTAC 1027

Qy      575 LeuValGlyArgGln--------------------------------------------- 579
            |||   |||||||||
Db     1028 CTCAAGGGCAGACAACTTCCAAAATCTCCGCCCATCACTTTAGATGATGGATTGGTGTAT 1087

Qy      580 -----------------------TrpArgAlaPhePheAla------LysAspAlaGly 589
                                   :::   :::|||              :::
Db     1088 GTAAGAAGGGTCGTAGTAACACCATGCAAAGTTTATTTTTGTGGTCCAGAGGTAAACGTA 1147

Qy      590 TyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysProIleIleLys 609
               :::   |||||    :::         :::|||||               ::::::
Db     1148 TCCAATCGGGTTCTGCGCAATTAT---------TCTGAAGACATAAATAACTTTCTTCGA 1198

Qy      610 -----------------------GluArgValHisPhePhe--------------Ala 616
                                   |||:::::|||                    |||
Db     1199 GTTTCTTTTGTTGATGAGGAGTGGGAGAAAATACATTCTACAGATTTATTACCAAGAGCT 1258

Qy      617 GluThrGlyIleThrPheArgProAspValPheLys--------------ThrArgSer 631
               ||||||         |||   |||:::::::::                |||:::
Db     1259 AGTACTGGAAATGGTACCAGGACAGACATCTATGAGAGGATCCTATCAACTCTTCGGAAT 1318

Qy      632 ValValProAlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeu 651
                  :::          :::                  |||   ||||||:::
Db     1319 GGCTTTATAATTGGTGATAAAAGATTTGAATTTCTTGCATTTTCATCGAGCCAGTTGCGG 1378

Qy      652 Asp---------Trp--------------------LeuLeuGlnLeuAspAsnAsnThr 661
            |||         |||                    |||         |||         |||
Db     1379 GATAATTCAGTGTGGATGTTTGCATCAAGACCTGGCCTTACTGCAAATGATATAAGAACA 1438

Qy      662 TrpGlnProHisLeuLysLeuPheSerArgIle--------------GlnLeuGlyLeu 676
            |||       :::         :::               :::||||||
Db     1439 TGGATGGGTGATTTTCGGCAAATCAGGAATGTCGCAAAATATGCTGCCAGACTTGGCCAA 1498

Qy      677 SerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAsp 696
            |||      :::      |||||                 :::      |||:::   :::
Db     1499 TCTTTTGGTTCCTCCAGAGAAACTTTG---------AGTGTTGGTAGGCATGAAGTTGAA 1549

Qy      697 LeuLeu---------------SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
            ::::::               |||   ||||||      :::   :::|||||||:::
Db     1550 GTTATTCCTGATGTAGCTTGTTCGCTTCATGGAACCAACTACATCTTTTCAGATGGAATT 1609

Qy      712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal 731
            |||:::::|||         |||   |||:::               ||||||
Db     1610 GGAAAAATATCTGCTGACTTTGCTCATAGAGTTACCATAAAATGTGGCCTTCAATATACT 1669

Qy      732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAsp 751
            ||||||:::   |||   |||:::|||         ||||||:::      :::|||
Db     1670 CCATCTTCTTTTCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGCTGTTGAT------ 1723

Qy      752 AspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAsp 771
                                          |||   |||   :::
Db     1724 --------------------------CCATATTCATCAATGAAG--------------- 1741

Qy      772 PheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791
                                   |||   :::|||         :::   :::|||
Db     1742 -----------------------TTATCTTTGAGAAAGAGCATGTTGAAATATGAATCA 1777

Qy      792 AlaGlyLeuAsnLeuGlnLeuLeu------------------ProValLeuGluAspArg 805
               :::   |||   :::|||                     |||         :::|||
Db     1778 AACAACATAAAGTTAGATGTCCTGGGATGGAGCAAATATCAGCCTTGTTATCTTAATCGT 1837

Qy      806 AlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGln 825
                  :::                  :::   |||   ::::::               |||
Db     1838 CAACTGGTTACTCTCTTGTCTACTCTTGGGGTGATAGATGATGTTTTAGAGCAGAAGCAA 1897
```

```
Qy    826 ArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpVal 845
Db   1898 AATGAAGCAGTAGATCAGCTTGATGCTATTTTGCATGATTCTTTGAAGGCACAGGAAGCT 1957

Qy    846 TyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAla 865
Db   1958 TTAGAATTGATGTCT--------------------------------------------- 1972

Qy    866 GlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspPro 885
Db   1973 ------CCTGGAGAGAATACTAATATTCTCAAGGAGATGCTAAATTGTGGTTATATGCCT 2026

Qy    886 LysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThr 905
Db   2027 GATGCTGAGCCGTTTCTTTCAATGATGCTGCAAACTTTCCGGGCTTCGAAGTTGCTCGAT 2086

Qy    906 LeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAsp 925
Db   2087 TTGCGGACTAGAACAAGAATATTTATTCCAAATGGTAGAACGATGATGGGATGCTTGGAT 2146

Qy    926 PheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSer------------ 941
Db   2147 GAATCCAGAACCCTGGAATATGGTCAGGTGTTTGTTCAGTTTTCTGGTGCTGGACGTAGA 2206

Qy    942 LysPheArgAspGluGluGluSerPhe------ThrLeuLeuSerAspCysAsp------ 957
Db   2207 CAGTTTTTTGAGGAATCGCATCCTTTTAATGACAGTGGATCCGCCAACTGTGATTTCATT 2266

Qy    958 ------------ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArg 973
Db   2267 CTCAAGGGAAACGTGGTCGTTGCAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTT 2326

Qy    974 ValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThr 993
Db   2327 TTAAGGGCTGTTGATGTTCCAGCTTTGCACCACATGGTTGATTGTGTTGTATTTCCCCAG 2386

Qy    994 LysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAla 1013
Db   2387 AAAGGAAAAAGACCTCATCCAAATGAATGTTCTGGAAGTGATTTGGATGGGGATATCTAC 2446

Qy   1014 TrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGlu 1033
Db   2447 TTTGTTTGTTGGGATCCAGACTTGATTCCGCCAAGGCAAGTCCAGTCGATGGATTATACT 2506

Qy   1034 ProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSer 1053
Db   2507 CCAGCACCCACCACACAGTTGGATCATGAT------------------------------ 2536

Qy   1054 HisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPhe 1073
Db   2537 ------------------------------GTCACAATTGAGGAAGTTGAAGAGTACTTC 2566

Qy   1074 HisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCys 1093
Db   2567 ------------------------------------ACCAACTAT--------------- 2575

Qy   1094 TyrIleAsnAsnSerValSer--------------------------------------- 1100
Db   2576 ATCATTAATGACAGTTTGGGAATCATAGCAAATGCGCACGTTGTATTTGCAGACAGAGAA 2635

Qy   1101 ---------------AsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVal 1115
Db   2636 CCTGATATGGCCATGAGCGATCCGTGCAAACAACTTGCTCAGCTCTTTTCAATTGCAGTT 2695

Qy   1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 1135
Db   2696 GACTTTCCAAAGACTGGCGTTCCAGCTGAAATACCATCT------CAGTTGCGCCCTAAA 2749

Qy   1136 LeuLeuGlyGlyAlaLeuSerLeuProAsp---ProMetTyrLysSerAspSerTrpLeu 1154
Db   2750 GAATACCCAGACTTCATGGAAAAGCCAGACAAGCCAACCTATCCCTCGGAAAAAGTTATT 2809

Qy   1155 GlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAla 1174
Db   2810 GGAAAG------------------------------------------------------ 2815

Qy   1175 IleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAsp 1194
Db   2816 ------------------------------CTTTTCCAGAAAGTGAAGAACAAA------ 2839

Qy   1195 GlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGluIleSer 1214
Db   2840 ---------------ACACCCCAGGCTAGCTCCATCGCGACCTTCACAAGGGATGTTGCC 2884

Qy   1215 AspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluVal 1234
Db   2885 AGGAAATCATATGATAGTGATATGATAGTCGATGGATTCGAAGATTACATTGACGAAGCT 2944

Qy   1235 ---------GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSerLys 1251
Db   2945 TTTTACTACAAAAGTGAATAC---------------GACAACAAGCTTGGTAATTTAATG 2989

Qy   1252 AspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrProGlu 1271
Db   2990 GACTACTACGGCATAAAAACTGAG-----------------GCTGAAATACTTAGTGGT 3031

Qy   1272 AlaMetAspLysSerGlyAlaAsnTyrAspSerLys------------------------ 1283
Db   3032 GGCATCATGAAAGCATCAAAAACTTTTGACCGCAGAAAAGATGCAGAGGCCATTGGTGTT 3091

Qy   1284 ValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp------ 1301
Db   3092 GCTGTGAGGTGTTTGAGG------------------AAGGAGGCAAGGGCCTGGTTCAAG 3133

Qy   1302 ---------------------AlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyr 1314
Db   3134 AGGCGTAGTGATATTGATGACATGTTAGCAAAGGCTTCGGCTTGGTACCATGTTACATAT 3193

Qy   1315 HisLysSer------------------------------------------ProLysPhe 1320
Db   3194 CATCATACATATTGGGGTCTCTACAATGAGGGGTTGAAAAGAGATCATTTCATTAGCTTT 3253

Qy   1321 ValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArgPro 1340
Db   3254 CCCTGGTGTGTTTATGACCAGCTGATCCAGATTAAGAAGGCAAAAGCTCGTAAAAGGCCG 3313
```

```
RESULT 5
AX204840
LOCUS       AX204840              3591 bp    DNA     linear   PAT 30-AUG-2001
DEFINITION  Sequence 2 from Patent WO0155407.
ACCESSION   AX204840
VERSION     AX204840.1  GI:15394184
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Beclin,C., Elmayan,T., Mourrain,P. and Vaucheret,H.
  TITLE     Novel sgs2 plant gene and use thereof
  JOURNAL   Patent: WO 0155407-A 2 02-AUG-2001;
            Rhobio (FR) ; INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (FR)
FEATURES             Location/Qualifiers
     source          1..3591
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
     CDS             1..3591
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC60174.1"
                     /db_xref="GI:15394185"
                     /db_xref="REMTREMBL:CAC60174"
                     /translation="MGSEGNMKKSVVTQVSIGGFGESTTAKQLTDYLEDEVGIVWRCR
```

```
                     LKTSWTPPGSYPNFEIADTSNIPSIDEYKKVEPHAFVHFAVFESAGRAMDAAGQCNLI
                     LDGQPLKVSLGPKNPYSLNQRRRTTVPYKLAGITLEIGTLVSRDDFFVSWRAEGVDFL
                     VDPFDNTCKFCFRKSTAFSFKDAVMHAVINCDYKLELLVRDIQTVRQYKTLHGFVLIL
                     QLASSPRVWYRTADDDIYDTVPGDLLDDDDPWIRTTDFTQVGAIGRCHSYRVLISPRY
                     ENKLRTALDYFRMRRVQEERVRWPPRIRNEPCFGEPVSDHFFCIHHKEGISFEIMFLV
                     NSVLHRGVFNQFQLTERFFDLLRNQPKDVNIASLKHLCTYKRPVFDAYKRLKLVQEWI
                     QKNPKLLGSHEQSEDISEIRRLVITPTRAYCLPPEVELSNRVLRRYKAVAERFLRVTF
                     MDESMQTINSNVLSYFVAPIVKDLTSSSFSQKTYVFKRVKSILTDGFKLCGRKYSFLA
                     FSANQLRDRSAWFFAEDGKTRVSDIKTWMGKFKDKNVAKCAARMGLCFSSTYATVDVM
                     PHEVDTEVPDIERNGYVFSDGIGTITPDLADEVMEKLKLDVHYSPCAYQIRYAGFKGV
                     VARWPSKSDGIRLALRDSMKKFFSKHTILEICSWTRFQPGFLNRQIITLLSVLGVPDE
                     IFWDMQESMLYKLNRILDDTDVAFEVLTASCAEQGNTAAIMLSAGFKPKTEPHLRGML
                     SSVRIAQLWGLREKSRIFVTSGRWLMGCLDEAGILEHGQCFIQVSKPSIENCFSKHGS
                     RFKETKTDLEVVKGYVAIAKNPCLHPGDVRILEAVDVPQLHHMYDCLIFPQKGDRPHT
                     NEASGSDLDGDLYFVAWDQKLIPPNRKSYPAMHYDAAEEKSLGRAVNHQDIIDFFARN
                     LANEQLGTICNAHVVHADRSEYGAMDEECLLLAELAATAVDFPKTGKIVSMPFHLKPK
                     LYPDFMGKEDYQTYKSNKILGRLYRRVKEVYDEDAEASSEESTDPSAIPYDAVLEIPG
                     FEDLIPEAWGHKCLYDGQLIGLLGQYKVQKEEEIVTGHIWSMPKYTSKKQGELKERLK
                     HSYNSLKKEFRKVFEETIPDHENLSEEEKNILYEKKASAWYHVTYHPEWVKKSLELQD
                     PDESSHAAMLSFAWIAADYLARIKIRSREMGSIDSAKPVDSLAKFLAQRL"
ORIGIN

Alignment Scores:
Pred. No.:                8.02e-06        Length:          3591
Score:                    328.00          Matches:         197
Percent Similarity:       37.17%          Conservative:    126
Best Local Similarity:    22.67%          Mismatches:      312
Query Match:              4.43%           Indels:          234
DB:                       6               Gaps:            42

US-09-913-878A-2 (1-1402) x AX204840 (1-3591)

Qy      535 PhePheGluIleLeuIleProSerProThrSerThrSerProSerValProProValVal 554
              ||         ::::::             ||||||:::|||   |||           |||
Db     1396 TTTGTTGCTCCTATTGTGAAGGATTTGACATCAAGTTCTTTCTCCCAGAAGACCTACGTT 1455

Qy      555 SerLysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSer 574
                |||:::      ::::::                  ||||||  |||
Db     1456 TTTAAAAGAGTGAAGAGCATA------------------TTAACCGATGGG---TTTAAA 1494

Qy      575 LeuValGlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeu 594
            |||   |||||::::::   :::|||   |||    |||
Db     1495 CTATGTGGTAGAAAATAC---AGTTTTCTAGCATTCTCAGCCAAT--------------- 1536

Qy      595 ArgGluPheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhe 614
                     |||||||||                                         |||
Db     1537 ---------CAACTGAGA-----------------------GACCGCTCTGCATGGTTC 1563

Qy      615 PheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValPro 634
            |||||||||   |||                        |||||||||
Db     1564 TTTGCTGAAGACGGG-----------------------AAAACACGT------------ 1587

Qy      635 AlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeu 654
                                                   ||||||   :::      |||:::
Db     1588 -------------------------------------GTGTCAGATATAAAAACATGGATG 1611

Qy      655 LeuGlnLeu---AspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGln 673
               :::      |||   |||            |||         :::|||:::
Db     1612 GGGAAGTTCAAAGACAAGAATGTG------------GCAAAATGTGCTGCTAGGATGGGC 1659

Qy      674 LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHis 693
            |||      |||   |||||||||   :::   :::   ||||||::::::
Db     1660 CTGTGCTTCTCCTCCACATATGCCACTGTAGATGTCATGCCTCACGAGGTT--------- 1710

Qy      694 LysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArg 713
              |||:::::::               |||   |||   :::||||||:::|||
Db     1711 GACACTGAGGTTCCAGATATTGAGAGAAATGGGTATGTTTTCTCTGACGGAATTGGTACA 1770

Qy      714 MetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal------ 731
            ::::::   :::|||      :::   :::   |||   |||   ||||||
Db     1771 ATCACACCTGACCTCGCTGACGAAGTAATGGAGAAACTTAAGTTG---GATGTGCACTAC 1827

Qy      732 ---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspVal 750
              |||   |||   |||   |||:::               ||||||:::
Db     1828 AGCCCTTGTGCTTATCAGATACGTTACGCAGGTTTCAAAGGGGTT--------------- 1872

Qy      751 AspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln--------------- 765
                                    :::   :::|||||||:::
Db     1873 -----------------------GTTGCTCGTTGGCCATCAAAAAGTGATGGAATCAGG 1908

Qy      766 ---------ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluVal 782
                     |||         |||   ||||||         ||||||:::
Db     1909 CTAGCCCTTCGAGACAGTATGAAGAAGTTCTTTTCCAAACAT---ACGATCTTGGAGATC 1965

Qy      783 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 802
               |||      :::      :::      ||||||   |||::::::   :::|||
Db     1966 TGTTCC---TGGACGAGGTTTCAACCTGGGTTCTTAAATCGGCAGATAATTACCCTTCTA 2022

Qy      803 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 822
                                                   :::   |||   :::
Db     2023 TCC-------------------------GTACTAGGTGTTCCGGATGAAATATTCTGG 2055

Qy      823 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArg 842
            |||:::|||               :::   ||||||||         ::::::
Db     2056 GATATGCAG---------GAATCCATGCTCTATAAACTGAACCGCATCCTTGATGATACA 2106

Qy      843 GlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValPro 862
                  :::|||   ::::::
Db     2107 GATGTGGCATTTGAAGTTCTCACGGCATCATGTGCT------------------------ 2142

Qy      863 PheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGly 882
                                    |||   |||         ::::::::::::|||
Db     2143 ---------------------GAACAGGGAAACACTGCAGCTATCATGCTTAGTGCAGGT 2181

Qy      883 PheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLys 902
            |||   ||||||   :::   :::|||:::   :::         ::::::      :::
Db     2182 TTCAAACCAAAAACCGAGCCGCATCTACGCGGGATGTTGTCTTCAGTCAGAATTGCACAA 2241

Qy      903 CysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMet 922
                  |||:::   |||         |||   |||            ::::::
Db     2242 CTCTGGGGTCTCAGAGAAAAATCTCGTATTTTTGTTACTTCAGGAAGGTGGCTAATGGGT 2301

Qy      923 IleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer------ 940
                  |||      |||:::||||||         :::   :::   |||
Db     2302 TGCCTAGACGAAGCAGGGATACTTGAACATGGCCAATGCTTTATTCAAGTCTCTAAACCG 2361

Qy      941 -----------------------------SerLysPheArgAspGluGluSerPhe 950
                                         |||:::|||::::::   :::
Db     2362 TCTATAGAAAATTGTTTCTCCAAACATGGGTCTCGTTTTAAGGAGACAAAGACAGATCTG 2421

Qy      951 ThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAsp 970
               ::::::               |||   :::|||::::::|||      |||   |||
Db     2422 GAAGTAGTTAAAGGCTAT---GTAGCCATTGCTAAGAATCCTTGTCTTCACCCAGGGGAT 2478

Qy      971 IleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIle 990
            ::::::   :::   ||||||      |||:::||||||   :::   |||   :::|||
Db     2479 GTAAGGATTTTAGAAGCTGTTGATGTACCCCAGCTGCATCACATGTATGACTGCCTTATT 2538

Qy      991 PheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGly 1010
                  |||||||||   |||         |||      ||||||   |||   ||||||
Db     2539 TTCCCTCAGAAAGGTGATAGGCCGCATACAAACGAAGCTTCTGGCAGTGACCTTGACGGG 2598

Qy     1011 AspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet 1030
            |||:::   :::|||   ||||||   :::::::::
Db     2599 GACCTGTACTTTGTGGCTTGGGATCAGAAACTCATC------------------------ 2634

Qy     1031 ProLeuGluProAspLeuSerArgTyr------LeuLysLysAspLysThrThrPheLys 1048
                  |||:::               |||         :::      |||         |||
Db     2635 ------CCTCCCAACAGGAAAAGCTATCCGGCCATGCATTATGATGCAGCTGAAGAGAAG 2688

Qy     1049 GlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMet 1068
              |||                   |||   |||         |||         |||
Db     2689 AGTTTA------------------GGCCGTGCTGTCAACCACCAGGACATAATCGATTTC 2730

Qy     1069 IleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsn--- 1087
                  ::::::               |||         ||||||         |||
```

```
Db        2731 TTTGCAAGAAAC-----------TTGGCGAATGAGCAGTTGGGCACAATTTGCAATGCA 2778

Qy        1088 ---------TyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAla 1104
                        :::   :::|||      |||                 ::::::
Db        2779 CACGTCGTTCATGCTGATAGAAGTGAGTAT---------GGAGCCATGGACGAAGAATGT 2829

Qy        1105 IleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPhe 1124
               ::::::|||:::   |||           |||||      |||   |||   :::
Db        2830 TTGCTACTGGCAGAACTAGCTGCCACTGCAGTTGATTTCCCAAAGACAGGGAAAATTGTG 2889

Qy        1125 AsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuPro 1144
               :::                     |||:::   :::|||            :::
Db        2890 TCAATGCCC------TTCCACCTAAAACCAAAACTCTACCCAGATTTCATGGGAAAAGAA 2943

Qy        1145 Asp---ProMetTyrLysSerAspSerTrpLeuGlyArg--------------------- 1156
               |||         |||||||||:::      |||||||||
Db        2944 GACTACCAAACTTACAAGTCGAACAAAATCTTGGGTCGGCTTTACAGACGGGTAAAAGAG 3003

Qy        1157 ---------------------------------------GlyGluProThrHisIle--- 1162
                                                      :::|||:::   |||
Db        3004 GTTTATGATGAAGATGCAGAAGCTTCCTCAGAAGAAAGCACAGACCCAAGTGCCATCCCT 3063

Qy        1163 ------------IleAspTyrLeuLysPheSerIleAlaArgPro--------------- 1173
                           ::::::         |||      |||
Db        3064 TATGACGCTGTTCTTGAAATACCGGGATTTGAAGATTTGATCCCTGAGGCATGGGGTCAC 3123

Qy        1174 ------AlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAsp 1191
                           |||   :::|||                  |||      |||:::
Db        3124 AAATGTTTGTACGACGGGCAACTCATTGGTCTTCTTGGGCAATACAAGGTGCAGAAAGAG 3183

Qy        1192 ThrGlu---AspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePhe 1210
                  |||            |||   |||                     ||||||
Db        3184 GAAGAGATTGTGACGGGTCACATCTGGTCC---------ATGCCCAAATACACAAGCAAG 3234

Qy        1211 LysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArg 1230
               |||:::               |||   |||                     ||||||:::
Db        3235 AAACAAGGCGAACTGAAAGAAAGA-----------------------CTGAAGCATTCT 3270

Qy        1231 IleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSer 1250
                        ::::::||||||:::   |||   |||      :::   :::   |||
Db        3271 TATAATTCCCTTAAAAAGGAGTTC---AGGAAAGTATTTGAGGAAACAATCCCTGACCAT 3327

Qy        1251 LysAspProTyrProValArgValAsnGlnValTyrGluLys------------TrpCys 1266
               ::::::                  |||   :::|||||||||            |||
Db        3328 GAAAATCTCAGCGAGGAGGAAAAAAACATCTTGTATGAGAAGAAGGCTTCAGCTTGGTAT 3387

Qy        1267 AlaIleThr------ProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysVal 1284
                  :::|||            ||||||   :::   ||||||
Db        3388 CATGTAACTTACCATCCCGAGTGGGTGAAGAAGTCT------------------------ 3423

Qy        1285 IleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrpAlaLeuLeu 1304
                        |||||||||            |||   :::   ::::::   |||   :::
Db        3424 ---------TTAGAGCTGCAA---------GATCCAGATGAGTCCAGTCATGCGGCGATG 3465

Qy        1305 ArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPheValTrpGlnMet 1324
                  :::                                    |||   |||   :::
Db        3466 CTGAGT--------------------------------------TTTGCTTGG---ATT 3483

Qy        1325 AlaGlyArgGlnLeuAlaTyrIleLys 1333
               |||         ||||||   ||||||
Db        3484 GCAGCTGATTATCTTGCAAGAATCAAA 3510
```

```
RESULT 6
NCB13B3/c
LOCUS       NCB13B3                 82584 bp    DNA     linear   PLN 06-MAR-2003
DEFINITION  Neurospora crassa DNA linkage group II BAC contig B13B3.
ACCESSION   BX284762
VERSION     BX284762.1  GI:28881246
KEYWORDS    .
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Schulte,U., Aign,V., Hoheisel,J., Brandt,P., Fartmann,B.,
            Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 82584)
  AUTHORS   German Neurospora genome,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2003) MIPS, Institut fuer Bioinformatik,
            GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,
            Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:
            G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute
            of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,
            E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT     this contig is an assembly of BAC 13B3 from 1 to: 11950 and BAC
            11C11 from 11951 to: 82584, strain OR74A, BAC clones are available
            at the Fungal Genetic Stock Center http://www.fgsc.net
            Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
            http://www.mwgdna.com
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of linkage groups II
            and V can be viewed at: http://mips.gsf.de/proj/neurospora.
FEATURES             Location/Qualifiers
     source          1. .82584
                     /organism="Neurospora crassa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5141"
                     /chromosome="6"
     repeat_region   3959. .3982
                     /note="24 bp tcatcatcatca tandem repeat"
     gene            5038. .6880
                     /gene="B13B3.010"
     CDS             join(5038. .5145,5216. .6880)
                     /gene="B13B3.010"
                     /note="similarity to clone mgc:8691, similar to amplified
                     in osteosarcoma, homo sapiens, TREMBL:BC000532_1"
                     /codon_start=1
                     /product="conserved hypothetical protein"
                     /protein_id="CAD70506.1"
                     /db_xref="GI:28881247"
                     /translation="MRRPSLALLALSSLPFGSARQPASFSIHQDLLAHPQFEVIFSDS
                     YVFEADAFALLEAANKTPKPTPASDGAHDGSTTRTDLTSAIRESATANADTDNGDESI
                     GGTSPLRETYELIAHPPMRYLCSIPIIAPPALNKTATELAKAEEAREVTRAYNKGWE
                     LMRGLENQCLHFVSGWWSYQYCYGKSIVQYHAVPNPKGGPPLRDKNSQEYILGTSLPP
                     SSHSQKGKQIEVPNNEQKQLSPPPNTELQAKDNQRYLVQRLDGGTICDLTGRPRTIEI
                     QYHCNPALSGDRIGWIKEVTTCAYLMVIHTPRLCADVAFLPPKETKAHPITCRQIITS
                     DEEALSFNQRRKNTIDSAAAAAAAVTTEDQKQGSESGSPEKLSYQGLTVAGIPIGARR
                     ILPSTHVLPLPRHLQQQRQEAQQGNLLEALTKAAFKADVFGDYGDDNNNNNNNHHPKA
                     GKGRKAAGAGKGQSGQKEMKKMRISERDIDKLGLDQQTLDALREEIRAAGLDPDRNLN
                     DEREAGGEIVWEFYADVSGDEDDGAVAEEGKEVFVWYEDEDEGGEPAEAGKDQKESKK
                     GGNGEGSGSGSEEGSKEEYYRDEL"
     exon            5038. .5145
                     /gene="B13B3.010"
                     /number=1
     intron          5146. .5215
                     /gene="B13B3.010"
                     /number=1
     exon            5216. .6880
                     /gene="B13B3.010"
                     /number=2
     repeat_region   7534. .7562
                     /note="29 bp poly(a) tandem repeat"
     gene            12653. .15824
                     /gene="B13B3.020"
     CDS             join(12653. .12719,12910. .13083,13261. .13438,13497. .14132,
                     14195. .15824)
                     /gene="B13B3.020"
                     /note="similarity to transcription activator, Aspergillus
                     oryzae, PIR:JC7252
                     Contains Fungal Zn(2)-Cys(6) binuclear cluster domain
                     [VCDKCRLRRIRCDYQSPCQPCIDAKLQC]"
                     /codon_start=1
                     /product="related to transcription activator"
                     /protein_id="CAD70507.1"
```

```
                /db_xref="GI:28881248"
                /translation="MMFMWMGAHGRRRGNLAILPGLQLRGSSSPEAQIVRLKTEPGTS
                TKQSPRPSSSSNATSAAGSSNRPVQPPKVCDKCRLRRIRCDYQSPCQPCIDAKLQCTR
                NHVPRKRGPRPGRGRLINRLREREAGKKEEQRPSSSGDPEVSDGARSNNTDIYPQNAS
                VSSTWASDVSTAPETTGLWANEHIFPSRGAFFYLIPRCVELFYQHLYPVMPVLYMPEI
                QRMDPRVCDLPEKNLLFSLSALTCFRMSGHSLGAEDSPEFWDQAGRFLLNDCLDVRKQ
                YDLYDNISLSTVVSSIFLASSFFETNQSTKAWVYLREALTFAQELGLEDESTYAGLSP
                EEALCRQRVFWLLYVMERSFAILRNKPLMLRRAPKLPMTLHAYESRGIHTGFLQLLGI
                YIPLRDSIIEAWTYGSNSTVDVNTYLALQNQLARPAGSSGGLSIPHPFFPSSPNAAVS
                SIQMPVSQSCSFSDLSLSSSSISSSPEDPVAPVVEPTASQTAELLVTQQWLRLIIWLS
                SLRQGYLSWAAENESMHFALPLTIARQTGLVLRSLMQIGGLEATGVAALEVNGMGIFE
                KIFEIGAWCMNVLDSYDKASLEGQFLTLPNFGRRHSAASGKNKQAARSATSSPADQVM
                DYEDGEDVDLLDVFMRALSATPTSRKQFAEPLYMWATTRPGGMRIGPGSVLHLDTGGG
                TGAFDMAQTQPGLAEQIPGLGLNQGLGLDHASGIGAAQGHDASSMMHQQQPGIWNAA
                SAYAHGQAPLPGSATGSGAGLSRPDGPGPGGGQGQGMEDVVMGDQASHQGQGFNIDPR
                LGAGSNVPRGLQHATFGRRPLRPNAPHLPSPLPGQFGGPPTSLYDVPGWSPIDPGGLT
                RARSRTMGMGQLRGLMAMVQGGPGTPGSSGPSVSQAGS"
exon            12653. .12719
                /gene="B13B3.020"
                /number=1
intron          12720. .12909
                /gene="B13B3.020"
                /number=1
exon            12910. .13083
                /gene="B13B3.020"
                /number=2
intron          13084. .13260
                /gene="B13B3.020"
                /number=2
exon            13261. .13438
                /gene="B13B3.020"
                /number=3
intron          13439. .13496
                /gene="B13B3.020"
                /number=3
exon            13497. .14132
                /gene="B13B3.020"
                /number=4
intron          14133. .14194
                /gene="B13B3.020"
                /number=4
exon            14195. .15824
                /gene="B13B3.020"
                /number=5
gene            16797. .17552
                /gene="B13B3.030"
gene            complement(join(16797. .16985,17226. .17552))
                /gene="B13B3.030"
CDS             complement(join(16797. .16985,17226. .17552))
                /gene="B13B3.030"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAD70508.1"
                /db_xref="GI:28881249"
                /translation="MHFIKSLLSLGLLSLLRYHAALALPTISDDTINDDTTINDDTLI
                PRVMIPKDPNMNFDPNNKPNVTKYSITDNPDHDVTLICPSPRALLSSGPPPEPGRFST
                TLGRWTTVSWTYQNAPTYKTLAEQAQMVLDQEKCRTTDATNTRPLVAGEAVYGHHDAW
                SVVLKGVVDCT"
exon            complement(16797. .16985)
                /gene="B13B3.030"
                /number=1
intron          complement(16986. .17225)
                /gene="B13B3.030"
                /number=1
repeat_region   17029. .17053
                /note="25 bp poly(t) tandem repeat"
exon            complement(17226. .17552)
                /gene="B13B3.030"
                /number=2
gene            18460. .24617
                /gene="B13B3.040"
gene            complement(join(18460. .24111,24165. .24617))
                /gene="B13B3.040"
CDS             complement(join(18460. .24111,24165. .24617))
                /gene="B13B3.040"
                /note="similarity to am-toxin synthetase (amt), alternaria
                alternata, TREMBL:AF184074_1
                Contains Putative AMP-binding domain signature
                [IQFTSGSTGKPK];ATP/GTP-binding site motif A (P-loop)
                [AAWSKGKT];Cysteine proteases inhibitors signature
                [QQVVQGALSFLGLN];Phosphopantetheine attachment site
                [AVGGDSLRAMRLIAAA]"
                /codon_start=1
                /product="related to AM-toxin synthetase (AMT)"
                /protein_id="CAD70509.1"
                /db_xref="GI:28881250"
                /translation="MQQVVQGALSFLGLNGQSSNATSHLFDTLGLETVSRPLSLPSQK
                NISLLDASLLAGLLLLQRERGEDDRVENFNWGRRTSDGAETSSQFSVKAVLGSEGLDQ
                TASVKNVLEVLKKVTGQDVTEENRTLFFNDGKEAPEKIKEEGYETAQRPNWKFQLELL
                ASEDGEQHDLTLQAYWFKNVPDALTSSQVTIRLNAFTELLGLIFSSPETTPISSLLAP
                LPCDLNQIWAFNAVVPPLIDRTMQDIISEQAAANPSKPAVSSWDGEFTYAEVESKSDL
                LARHLVSRGVKLGGVVPLMFEKSRWTIVALLAVMKAGSAFALTDPTSQPEGRLRVMVE
                QTGGDIIVASASQTELAEVCLPSSSGGSQGQVITVSDDLFTTLSQTPELTSVTLPVIP
                TQTTPLYIQFTSGSTGKPKGVVISHANYTSGALPRAEAVGYKSSSRCFEFASYAFDVS
                IDCMLCTLAVGGTICIPSDADRMNDLGGAILKSGANMAHMTPSVARVLDPEVIKGLDV
                LGLGGEAVSAADAAAWSKGKTSVIVAYGPSECTVGCTVNNVFARDEEAKKKGFTTGNI
                GTGVGGCGWVVDPNDHDRLVPVGAVGELLMEGPVVGLGYLGEEEKTKEVFIEDPKWLV
                AGHGDAHPGRKGRLYKTGDLVRYDADGSGAFVFVGRKDAQVKLRGQRVELVEIEHHLR
                AKLPSGVKIAAEVIKPSGGDPTLVAFLAETSSTGISAKDALQSVVEGEVSFSPELQAA
                LDGIEEVLGIDLPRYMVPNAYIPLREMPSLVSGKIDRKTLRAMGAAMTREQVSGSAKK
                ARQGGEGGAPTTEMEIALHKVWKQLLGTEQEIAAGDSFFAVGGDSLRAMRLIAAARAE
                GVKLTVADVFRYPVLRDMAVVATKVDTQDAGGKDNNEIAPFSLLESDWAIEDVKAEAS
                KNCNIDKDALEDVYPCTPLQEALMALSAKVKEAYVAQRVLKMEGQAAADKLKAAFEAI
                NADCAILRTRIIHSPRGLMQVVVKEPLHWHSATTLAEYLETDRNEEMDLGKPLVRYGL
                IQDGDAYHIVLTMHHALYDGWSMPLVVDRVNQAYQGLVPRKPAAEFKHFINYLNNTLD
                RAACDAYWKDQLAGATGVQFPALPFEGYQTRADSLWELYVSLDGRKPPSSPNTTITLA
                TIVRAAWALVASQYTAGNNDIVFGETLTGRNAPIVGADEIEGPMITTVPVRVCIDREQ
                TVAEYLQQVAEQMITQIPYEHAGLQHIRRLSDDALQACELRTGIVLHPAAGEVGVTEN
                TPANGLVPAGDSEAAQEALKFNTYALMLVCSLEADGFLVMASFDSKTVEMDVVKKALD
                QLQVVVRQLCEDGCGEKKVGQLECLTEKDHEEVKKLAGVKNLTAESDPFVGGVTAEDV
                EGVWIVDAADTARLVPVCAAGELIVETSKELDAPVTLLAETPAWYEGNKKLYKTGQLA
                KFKVDGDSASLCILKASEPKTEVAKKKVASSVTPSAVAATSAKQKTLRKIWAKLLKVK
                EDSIYLGDSFFNRGGDSITAMKLVSEARQQGLKISVAQVFANRTLFDMANVMQTTGTV
                VDASEKVETEYKPFSLLEENNKASIMEVVKHSLADKSWTIADILPTRPLQEIAVRGTV
                ELPRFSLRYELMHFNGQVDKAALFRACHEVVARNEILRTVFVRDASNTCYGAVIDNLI
                TSIAEYEIDDHADVKDFAAQVCRLDSQTRMPYGSSFVKWFFITNGTSQCSLAFRLSHA
                QYDEICLPIMLHQFHQLYLNSPESSIPPSYPFSHFVAHTLRDSIPAAIPYWRNLLSGS
                SGITRFRPSTPITSRKHFAIHRAVDISSRPRDVTIATLPSAAWALALARHVKSPDVLF

Alignment Scores:
Pred. No.:              0.00462       Length:         82584
Score:                  306.00        Matches:        274
Percent Similarity:     33.15%        Conservative:   143
Best Local Similarity:  21.78%        Mismatches:     440
Query Match:            4.14%         Indels:         402
DB:                     8             Gaps:           57

US-09-913-878A-2 (1-1402) x NCB13B3 (1-82584)

Qy        132 ProProProAsnAsnThrProGlyArgThrPheGlyArgThrLeuSerGlyProSerGly 151
              |||   ||||||::::::                        :::   :::|||      |||
Db      45061 CCTCATCCCAACTCTTCAAGGAAGAACCCAGCGCACGACAGCTTTGCAGGACGGCCCGGG 45002

Qy        152 LeuSerArgPro------------------------------------------Thr 156
              |||      |||                                             |||
Db      45001 CTGCGTTCGCCGTGTCCACTCCCGCCTGTCGTCTCGTTTCTCTGTTGTTGTTCACGGACG 44942

Qy        157 SerThrAsnThrLysArgLysAspGluProAlaAsnValThrPheAlaAspPro------ 174
                       :::   |||                    :::|||   |||   |||
Db      44941 CACCGACGTTCAACGCGCCTGTCGTCTCGTTTCTCCGTGTTGTTCACAGACACACTGACG 44882

Qy        175 ------------------------ProLysArgSerLeuThrArgSerAlaThrGlyPro 186
                                      |||         |||:::||||||:::
Db      44881 GGTTCAACCCCTCTGTTGTTCCGTCCGCCCACCAAGCTTTCAAGGTCCTCAATTTGTCTC 44822

Qy        187 ProIleHisGlyAlaAlaIleProLeuLysPheProAspProValAsnThrGlySerLys 206
```

```
Db     44821 AAGTTGGAG-----------CCTTTAGCACTCACTCACTTGGTCAACCTTTGGTGAGCT 44774

Qy       207 ArgProSerLeuGluSerGluAsnLeuAsnGlnCys----------------------- 218

Db     44773 TCTCCCCTTCTCTCAAACATGTTGATCCAAAAATGCTGACACCCACATAGCCATCCTAAA 44714

Qy       219 ------ThrLysArgAlaLysGlyLysLeuSerAsp----------------------- 228

Db     44713 TAGCTGGTCGACAGCTCTATGCGCAAATTATCCTCTCGCCCGCCATGGAGGTCTACATCA 44654

Qy       228 ------------------------------------------------------------ 228

Db     44653 GGAACCTTCCTCAAGGCCTTTCCGAGAGAAGCTTGACAGCTCATCTGAAGCCGTTCATGG 44594

Qy       229 -----------------AsnValAlaAlaAlaAlaAlaProProValProIleAlaSer 242

Db     44593 ACAACCTTGGCATCCGTCACTACGACTGCAATAAACGGTCCAACCAACCCTTTGGGAAAG 44534

Qy       243 AlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAsp-ProThrAlaThrGl 262

Db     44533 TCATCTTTATCACTGCCAGGGAAGCGCTGAACTTCATCGCGAAGCACCCTCGTCTTTATA 44474

Qy       262 yHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyrGlySe 282

Db     44473 TCATGAATCAGCAGGCTGAATGCAGGGT---------CAGCAAGCCCCAGCCAAAGGATC 44423

Qy       282 rSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSer-PheGluAlaP 302

Db     44422 TGAAGAACTTGGTCAA---CGGCAT--------------------TCTTTACGAGGCTG 44387

Qy       302 roProSerGlnProArgGluLys-----------------------------ArgPro- 311

Db     44386 AAGAAAAGGAAAGGCGAAAGAAGGCGTGGGAGGAGAATACGAGCAACAATGCCCGGCCTG 44327

Qy       312 -------------------ValAspAlaThrValPheGluAlaGlyHis---------- 321

Db     44326 CACGCACGCCAACTTTAGCACTTGATGCCGACGTTCTTGACTGCGGACATTGGTCCTACA 44267

Qy       322 ----------------LeuIleGluSerProSerLysGlyArgThrThrLysSerHisI 336

Db     44266 TCGACGGCAACTTGACCTTCATCACCGAGTGGTCGTCCTCCCTGCGCGTCTCAGCAAAGT 44207

Qy       336 leAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThrTyrTyr- 355

Db     44206 TCGCCAAGCACGACCTCGTCATTACCTCTTTGGCTGAATCTCAGGTATCTATCCCATATC 44147

Qy       356 -----------GluSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerA 372

Db     44146 GATATATCCACGAACTCGTTTGGTCGAATGGTGGACATGTCGCTGTC------------ 44100

Qy       372 rgSerAsnGlyLeuAlaArgSerGluGluSerAlaArgSerGlnValGlnValHisAlaP 392

Db     44099 ------------------------------------------ACTCTCGCACACGCCC 44084

Qy       392 roVal---ValAlaAlaArgLeuArgAsnIleTrpProLysPheProLysTrpLeuHis- 410

Db     44083 CTACGTTTCTCTCGCCGCCAATTCCATTACTTCGGCCCGGGCAACCCAAGAGGATGCGCC 44024

Qy       411 --GluAlaProLeuAlaValAlaTrpGluValThrArg------LeuPheMetHisCysL 428

Db     44023 TCGAGGCATGTGATAGTAACCACGCCAAAGTATCAAAATTCTGCCTCGTCTATCATTTCA 43964

Qy       428 ysValAspLeuGluAspGluSerLeu-----------------------GlyLeuLysT 440

Db     43963 AGGTT------TCGGACAAACATTTAAAACAGAACCAGGGTAGCTACAGAGGGAGTGACT 43910

Qy       440 yrAspProSerTrpSerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgL 460

Db     43909 TCCGTACCGCCATTAATGGACTGAATGAGCAAGAAATGTTCTGGGTCACGCATTATACAT 43850

Qy       460 euAspAlaPheArgGlyLysProPheProGluLysProProAsnAspValPheValThrA 480

Db     43849 TC-----------------GCCATTCAAACCACATCTCCGCAAGATGCATATGCAGTTG 43808

Qy       480 laMetThrGlyAsnPheGluSerLysGlySerAlaValValLeuSerAlaValLeuAspT 500

Db     43807 CAGTTGACAGG-------------------------CTTAGGGCGGCTCTTCTGGAGT 43775

Qy       500 yrAsnProAspAsnSer------------------------------------------ 505

Db     43774 ACGAAGGTAGAGATACGTTGCCTTACAGCCTGCTCTTCAATCTCCAAGCGCTGGTGTACA 43715

Qy       506 --------------ProThrAlaProLeuTyrLeuValLysLeuLysProLeuMetPheG 521

Db     43714 TGTCGTATCTTCATCCGACAACCGTGCTAAAGCTAGCCAAGCGCTTAGCTGATATGTTCG 43655

Qy       521 luGlnGlyCysArgLeuThrArgArgPheGly-ProAspArgPhePheGluIleLeuIle 540

Db     43654 AAACGGCCCGACGCTCAGGTCAGCGGCAGGATCCCATTTCGG-----------TTGATG 43607

Qy       541 ProSerProThrSerThrSerProSerValProProValValSerLysGlnProGlyAla 560

Db     43606 CCTTCAAGGACCTCTTCAAAACCATCGACTGG-CCG------TCGCCGAGCCTCTCGAGC 43554

Qy       561 ValGluGluValIleGlnTrpLeuThrMetGly------GlnHis---SerLeuValGly 577

Db     43553 GCAGAGGAGCTCGCTCAGTTTGAGGTCGAAGGGATTATTGAACACTTGAAGAAAACGGAG 43494

Qy       578 ArgGlnTrpArgAlaPhePheAlaLysAspAlaGly----------------------- 589

Db     43493 AAACGCATGCGTGAGGGCTACACCCTGCGGTTAAACGAAGAGATTCCTCCCGGTCTTACC 43434

Qy       590 ------TyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysProIle 607

Db     43433 AAGATCTATCGGGCACTTGTCACACCAACCCGAATCGAACTTCACGGGCCTGAACTGGAA 43374

Qy       608 IleLysGluArgVal--------------------------HisPhe------------ 614

Db     43373 GCCAAGAACAGGATCCTTCGAAAGTTTCCAGAACACCAGGACCACTTCCTTCGAGTCCAG 43314

Qy       615 PheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValPro 634

Db     43313 TTCGCCGAAGAGGAT---------GGACAGGACCTCTTTTTCAACTCGGCGGTT------ 43269

Qy       635 AlaGluGluProValGluGlnArgThrGluPheLys----------------------- 646

Db     43268 TCCATGGACGCCATATATCAGAGA------TTTAAAGACGTCTTGACCAATGGCATCTCA 43215

Qy       647 ------------ValSerGlnMetLeuAsp---------------------------- 652

Db     43214 GTTGGTGGACGGGTTTACAGGTTTCTCGGCTTCTCCCATTCCTCCCTCCGAGCACATAGC 43155

Qy       653 ---TrpLeu--------------------------LeuGlnLeuAspAsnAsnThrTrp 662

Db     43154 CTTTGGCTAGCGGCGCCTTTCATATACGACGGTAAATTGCAATTAGCTAGTAACATCATT 43095

Qy       663 Gln-------------------------ProHisLeuLysLeuPheSerArgIleGln 673

Db     43094 GAGGATTTGGGCGATTTTCGCAATATCATGTCGCCAGCGCGACGTGCGGCACGCATAGGC 43035

Qy       674 LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHis 693

Db     43034 CAGGCCTTTTCAGAGACACCATATTCCGTTTCGCTCTACGATCACGGTATT--------- 42984

Qy       694 LysThrAspLeuLeu---------SerProSerGlyThrGlyGluValMetAsnAspGly 710

Db     42983 ------GATGTGATAAGGCAAAGGGACGTCAAACGCAACGAACGCGTGTTCAGTGATGGC 42930

Qy       711 ValGlyArgMetSerArgSerValAlaLysArgIle---ArgAspValLeuGlyLeuGly 729

Db     42929 GTTGGGATAATTTCGCAGGGTGCCCTCGAGGTGATTCATCGCGAGATTCCCGAATCCAAG 42870

Qy       730 AspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp 749

Db     42869 GGCTATCCCAATTGCTTACAGGTTCGATGGGCTGGAGCAAAGGGCATGTTGGCTCTCGAT 42810
```

```
Qy        750 ValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGlu 769
                        ||||||    :::        |||           |||      |||
Db      42809 GCCCGGCTCACTGGGAGGCAAATCTGTATCAGG------GATTCCATGGAGAAG------ 42762

Qy        770 CysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeu 789
                    |||      :::    :::      ||||||:::       :::||||||:::
Db      42761 ------TTCCGCAGCCGTGATGAGGAACATCTCGAAATATGCGACATGGCTTCCAAGCCA 42708

Qy        790 LysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys 809
                          ||||||   |||::::::   :::|||||||||
Db      42707 ATTCCTCTGATGCTGAACCGTCAGATGATCAAAATTCTCGAGGAC--------------- 42663

Qy        810 ValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSer 829

Db      42663 ------------------------------------------------------------ 42663

Qy        830 GluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyr 849
                                 :::     |||          ||||||      |||
Db      42662 ---------------ATGCGCGCCCCTGCC---------CAGTGGTTCTTGGAGCTTCAA 42627

Qy        850 SerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAsp 869
                             |||                       |||    :::   |||
Db      42626 GAAAAGGAACTCCAACGT-----------------------CTGCGGGCAATCACGGAT 42591

Qy        870 SerGlnGluGluThrLeuAsnPheLeu--------------------------------- 878
              :::   :::           ||||||
Db      42590 AACGTCCAGAACGTGGCAACCTTTCTCAAGCTGCAGTGTGTCGGAGATTCTGTACACCTC 42531

Qy        879 ---------------------------MetAsnSerGlyPheAspProLysLysGlnLys 889
                                         ||||||        |||   ::::::  :::
Db      42530 AGCCAGTTTCTCAAAGACCTTGACAAAATGAAC------ATCGACTACCGGAGAGACCAG 42477

Qy        890 TyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLys 909
              :::|||:::   |||            :::    |||:::         ||||||   |||
Db      42476 TTTCTTCGCGGTATTGTCGAAGCTGTCGTGCTTAGAGAGCTGCGCTTGCTCAAGCACAAG 42417

Qy        910 LeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyVal 929
                 |||   |||               ::::::   :::   |||         |||:::
Db      42416 GCTAGAATCCCGGTTCCCTATGGCGTTACTCTGTTTGGTGTCATGGACGAGACAGGCCTA 42357

Qy        930 LeuGluGluAsnGluValHisValGlyPheSerSer---------LysPheArgAspGlu 946
               |||   |||   ||||||:::|||   |||   :::         :::|||:::|||
Db      42356 TTGCGCGAAGGCGAAGTATACGTGACTTTTGAGACCGTGGACGGTCGCTTCAAGGACCCT 42297

Qy        947 GluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHis 966
                 :::                              |||:::|||   ||||||||||||
Db      42296 CCCACTGCT------------------GGCCCCGTTGTGGTGACCAGGTCTCCAGCCTTG 42255

Qy        967 PheProSerAspIleGlnArgValArgAlaValPheLysPro---------GluLeuHis 983
                 |||   |||||||||         |||                     |||
Db      42254 CACCCTGGAGACATCCAGATTGCCCACAATGCCATTCCACCGGCAGGGCACCCTCTCAGA 42195

Qy        984 SerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeu 1003
                 ||||||:::   |||:::||||||      |||:::      |||      :::|||
Db      42194 GAACTGAAAAACTGCATCGTGTTCAGTCAAAACGGTGAACGTGACCTGCCTAGTCAGCTG 42135

Qy       1004 SerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleVal--- 1022
              ||||||||||||   ||||||||||      |||   ||||||      ||||||
Db      42134 TCGGGCGGTGACTTGGATGGCGACACTTTCAACGTGATCTGGGATCAAAGCATCGTTGCC 42075

Qy       1023 ------AspGlyPheValAsnAlaGluMetPro---LeuGluPro---AspLeuSerArg 1038
                    |||   |||:::      :::|||   |||   :::||||||   |||:::|||
Db      42074 ACGCTCAGAACTTTCGCGGCGGCCGATTATCCACGTGTCGAGCCCTTGAAACTGAACCGC 42015

Qy       1039 TyrLeuLys---LysAspLysThrThrPhe------LysGlnLeuMetAlaSerHisGly 1055
                 ::::::   ||||||         |||      :::   :::   |||   |||
Db      42014 GAGGTCGAGTCAAAGGACATGGCTGACTTCTTCGTGGAGTTTATGAAAGCCGATCACCTG 41955

Qy       1056 ThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPhe 1075
                       |||   :::               |||
Db      41954 GGGGTAATTGCGGTTCGGCACATGATCCTCGCGGAC------------------------ 41919

Qy       1076 AlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIle 1095
                                                         ||||||
Db      41918 ------------------------------------------GAACGC------------ 41913

Qy       1096 AsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVal 1115
              |||         :::         :::   |||::::::|||               |||
Db      41912 AATGAGGGAACGCTGGATGCAGACTGTCTCAAGCTGGCAGCATTGCACTCTAAAGCGGTT 41853

Qy       1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 1135
              |||   ||||||   ||||||                         :::|||   |||
Db      41852 GACTTCTCCAAATCGGGGATT---------CACGTGGATATCACCGAGTTACCAAGG--- 41805

Qy       1136 LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp 1151
                                          |||||||||:::   |||
Db      41804 --------------------------CCACCAATGTATCGGCCTGAT 41784
```

```
RESULT 7
AY029284
LOCUS       AY029284                6847 bp    DNA     linear   PLN 28-DEC-2001
DEFINITION  Neurospora crassa suppressor of ascus dominance (sad-1) gene,
            complete cds.
ACCESSION   AY029284
VERSION     AY029284.2  GI:13699900
KEYWORDS    .
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (bases 1 to 6847)
  AUTHORS   Shiu,P.K., Raju,N.B., Zickler,D. and Metzenberg,R.L.
  TITLE     Meiotic Silencing by Unpaired DNA
  JOURNAL   Cell 107 (7), 905-916 (2001)
   PUBMED   11779466
REFERENCE   2  (bases 1 to 6847)
  AUTHORS   Shiu,P.K.T., Raju,N.B. and Metzenberg,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2001) Biological Sciences, Stanford University,
            385 Serra Mall, Stanford, CA 94305-5020, USA
REFERENCE   3  (bases 1 to 6847)
  AUTHORS   Shiu,P.K.T., Raju,N.B. and Metzenberg,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2001) Biological Sciences, Stanford University,
            385 Serra Mall, Stanford, CA 94305-5020, USA
  REMARK    Sequence update by submitter
COMMENT     On Apr 20, 2001 this sequence version replaced gi:13641450.
FEATURES             Location/Qualifiers
     source          1..6847
                     /organism="Neurospora crassa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5141"
                     /chromosome="I"
                     /map="between leu-3 and cyt-1"
     gene            997..>6375
                     /gene="sad-1"
     CAAT_signal     997..1001
                     /gene="sad-1"
     TATA_signal     1119..1125
                     /gene="sad-1"
     mRNA            join(1204..4542,4601..>6375)
                     /gene="sad-1"
                     /product="suppressor of ascus dominance"
     CDS             join(1401..4542,4601..6375)
                     /gene="sad-1"
                     /note="SAD-1; RNA-dependent RNA polymerase; required for
                     meiotic gene silencing of unpaired DNA"
                     /codon_start=1
                     /product="suppressor of ascus dominance"
                     /protein_id="AAK31733.1"
                     /db_xref="GI:13641451"
                     /translation="MSGIPNWVEQAQHAILVSPDRSPKDQDARPRHPSRRRSSQQGHR
                     ENTPTPRPPSRPTSRPSSSSPKPPPRAFRQFSRGKPASFASLPSLTQPVFISREWKCA
                     RTAAIRVKDVPPRAKLQDLHDLFSEYGHISYIELDEDRQDVSDRGRRALIRFEPPPLS
                     TDFLCRGICRLRIGLVDHWLPLETVDPKREDRKERTIKTQLGNTCPQTLYVFPSTLSF
```

```
                GLLVQPAVFMKKQSVQTLSSDSFLRLEFDFKRMRLVIKFSLHYQRELSGFRRQNEHYK
                LHIKFGVIKELCRTMVGEEHRQALVITLRDPPVAYRKRDVSKTFGEDRLTWSENDLWE
                RVVDISPGLDVSKNPVSLAENHQYIDLGRWLTYWIELDQQSTRVWDQVQQYLLDWNLR
                TKLTVFPEPLPNQKPKVWDFLDDRYGHDIQQVSSRSWSNDFSLLAAPPRISLPFDVRY
                QLEVCISQGIINEHNIDRPFLEKLMEFCDNDNSFGKDRARLILEYVADEYAGKRIFDP
                MELFKDHAALAYFPTSFMIPNHCAWVRRVTITPTRIYFSTPCVEPTNRVIRQWKHAQD
                YFIRIQFTDEVLEGRIKSGEAELPLFLRVYRVLEKGVAMGPWHWKFLAFGNSQIREAG
                AFMFCEQSNLTGDMMRAWMGRFSHIKVIAKYAARLGQCFSTTRLVPGIPAPRIVTIPD
                VEKDGFCFTDGVGKISPLLAKIVAHDWSIDPPPSAYQFRMGGCKGVLVTWPDVKGMEV
                HIRKSQEKFVAEFNGLEVVRCSQFSTATLNRQIIAVLSSLGVPDQVFVDMMEQQLSDF
                NAAMEDKQKATAILKTFIDENHMTPIIAEMLAYGFMGSQEPFVRTLLQLWRSWSIKTL
                KEKARLNVEKSAFVLGCVDETGTLKGHMKVIEDWKDVSSEKLPQIFLQIPDDVNGGYR
                VITGTCVVGRNPSLHPGDIRVVEAVDVPALRHLRDVVVFPLTGDRDVPSMCSGGDLDG
                DDFFVIWDPLLIPKERSHPPMISEPIAGKELATEPTVNNLITFFVLYMKYNNLPLIAH
                AHLATADAEVEGVKSPKCLELASLHSMAVDYVKTGVAAEFPRRLDPKTWPHFMEKNRH
                TYHSVTALGKLYDMVKRETFDMKENYQLPFDNRILKHTKCRALRDGTLAKARRIKSQY
                DTAMRRVMCQLEIATEFEVWTAFVMSKPRVGSDYKLQDNVGRESSALKQHFKDQCKKE
                AGGDLLSFVSAMYRVTYEEVRIALFEAKQPHVRPDGRLGTRKITPKTMPLVSFPWLFW
                DKLGELARAGAVLQRRLDDGSEDMDLLSDVPLVSQRRRGKHNASGSSDFMDEHGDPLS
                YTRTSDGKIIHYGQILNLFSHDDEDGDERNDARDRNSSSEDSTHVSNSSKSSSNLSPV
                AEEDLLTFESPPASPGVTPASPQVDLLGPTLTLTKAELAIGLAGPVATTYTPPPTDEE
                TVICHGHSNRASHLSASSSSLDPQPLIEETSGENEDLLLDIYSASPPRADGARLASAV
                GSDVPVKGGPPPVWMEQVIRMGHRIPTPPLDSIAIPNVIGVPQFESPLTGTVVASAIQ
                DPFVSPSPAVATPATSGSGGGWGGGGGGGGGYIGIGGPINKELVMGLDGGIKEEDRE
                ETDEEAEEVELEIDEDPIAVRYAQMAAL"
ORIGIN

Alignment Scores:
Pred. No.:              0.000373        Length:         6847
Score:                  299.00          Matches:        302
Percent Similarity:     33.44%          Conservative:   206
Best Local Similarity:  19.88%          Mismatches:     510
Query Match:            4.04%           Indels:         506
DB:                     8               Gaps:           67

US-09-913-878A-2 (1-1402) x AY029284 (1-6847)

Qy        89 AsnTrpValProLysAla---HisAla---------AspProAspThrLeuProTrpSer 104
Db      1416 AACTGGGTAGAGCAGGCGCAGCATGCCATTCTTGTGTCGCCGGACAGAAGCCCCAAAGAT 1475

Qy       105 LysGluPro------------ProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGln 120
Db      1476 CAGGACGCTCGTCCGAGGCATCCTAGCCGCCGCAGATCCAGCCAGCAGGGACATCGAGAG 1535

Qy       121 ThrValLeuLeuGluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArg 140
Db      1536 --------------------AACACACCGACACCTCGACCACCCTCGCGACCA------ 1568

Qy       141 ThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThr 160
Db      1569 --------------------------------ACCTCGCGACCCTCCTCCAGCTCTCCC 1595

Qy       161 LysArgLysAspGluProAlaAsnValThrPheAlaAspProProLysArgSerLeuThr 180
Db      1596 AAG-----------------------------------CCTCCCCCTCGCGCTTTTCGG 1619

Qy       181 ArgSerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspPro 200
Db      1620 CAGTTCTCACGGGGCAAACCAGCCTCATTCGCATCCCTCCCC---TCTCTGACACAACCA 1676

Qy       201 ValAsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLys 220
Db      1677 GTCTTCATTTCCAGGGAGTGG--------------------------AAGTGCGCCAGG 1709

Qy       221 ArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValPro--- 239
Db      1710 ACCGCTGCCATTCGCGTCAAGGAT---GTGCCGCCTAGAGCTAA-ACTCCAGGACCTTCA 1765

Qy       239 ------------------------------------------------------------ 239
Db      1766 CGATCTCTTCTCTGAGTATGGTCATATTAGCTACATTGAACTTGATGAGGACCGTCAAGA 1825

Qy       240 -------------------------------------IleAlaSerAlaLeuAspLys 246
Db      1826 TGTTTCTGATCGTGGCCGGAGAGCTCTGATTCGCTTTGAGCCACCTCCGCTATCTACCGA 1885

Qy       247 ValProThrArgArgHisAlaAsnThrArgAsp--------------------------- 257
Db      1886 CTTTCTCTGCCGAGGCATCTGCAGACTCAGGATTGGACTTGTTGACCACTGGCTGCCTCT 1945

Qy       258 ------------ProThrAlaThrGlyHisArgArgAlaAspGlnValAsp--------- 270
Db      1946 CGAGACAGTGGACCCAAAGCGCGAGGATCGCAAAGAACGTACCATCAAGACTCAGCTCGG 2005

Qy       271 -----------------SerPheAspThrSerGlnGlyThrSerTyrGlySerSerVal 284
Db      2006 CAACACTTGCCCGCAAACTCTATACGTCTTTCCCTCAACACTGTCTTT---CGGTCTCTT 2062

Qy       285 PheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGluAlaProProSer 304
Db      2063 AGTCCAGCCTGCGGTATTCATGAAGAAACAATCCGTTCAAA------------CCCTCT 2109

Qy       305 GlnProArgGluLysArgProValAspAlaThrValPheGluAlaGlyHisLeuIleGlu 324
Db      2110 CGTCCGACTCTTTCCTGAGACTTGAGTTCGATTTCAAGCGCATGCGACTGGTAATTAAAT 2169

Qy       325 SerProSerLysGlyArgThrThrLysSerHisIle------------AspAsnGlnPro 340
Db      2170 TCTCCCTCC---------ACTACCAACGCGAACTTAGTGGGTTTCGCAGACAAAATGAGC 2220

Qy       341 LeuSerSerSerSerGlnGlyGluThrSerPheSerThrTyrTyrGluSerPheProSer 360
Db      2221 ATTACAAGCTCCACATTAAGTTTGGAGTCATCAAAGAGTTATGCAGAACCATGG------ 2274

Qy       361 SerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLeuAlaArgSerGlu 380
Db      2274 ------------------------------------------------------------ 2274

Qy       381 GluSerAlaArgSerGlnValGlnValHisAla---------------ProValValAla 395
Db      2275 ---TCGGCGAGGAGCATCGTCAAGCTTTG-GTCATAACTCTCCGGGACCCGCCGGTTGCG 2330

Qy       396 AlaArgLeuArgAsnIle-----------------------------TrpProLysPhe 405
Db      2331 TACAGGAAGAGGGATGTCTCAAAGACATTTGGCGAGGATCGGCTGACTTGGAGCGAGAAT 2390

Qy       406 ProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArg--------- 422
Db      2391 GACCTCTGGGAGAGAGTGGTCGACATTAGCCCTGGACTCGATGTCAGCAAAAACCCAGTT 2450

Qy       423 -----------------------------------LeuPheMetHisCysLysValAsp 430
Db      2451 TCGCTCGCCGAGAACCACCAATATATTGACCTGGGACGGTGGCTTACCTATTGGATTGAA 2510

Qy       431 LeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrpSerThrAlaArgAspVal 450
Db      2511 CTGGATCAACAATCAACAAGAGTC---------------TGGGATCAAGTTCAACAATAT 2555

Qy       451 ThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArgGlyLysProPhePro--- 469
Db      2556 TTGCTCGATTGGAATCTCAGGACTAAGTTGACTGTTTTC---CCCGAGCCGCTTCCCAAC 2612

Qy       470 GluLysProPro-AsnAspValPhe---------ValThrAlaMetThrGlyAsnPheGl 486
Db      2613 CAAAAACCAAAGGTATGGGATTTTCTGGATGATAGATACGGCCATGACATTCAACAAGTC 2672

Qy       486 uSerLysGlySerAlaValVal-------------------------------------- 493
Db      2673 TCCTCACGGTCCTGGAGCAACGACTTTTCTCTTTTAGCCGCGCCGCCCCGTATTTCTCTA 2732

Qy       494 ----------------LeuSerAlaValLeuAspTyrAsnProAspAsnSerProThrAl 508
Db      2733 CCGTTTGATGTTCGGTATCAGCTTGAAGTCTGCATATCGCAAGGAATCATCAACGAGCAC 2792

Qy       508 aProLeuTyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrAr 528
```

```
Db   2793 AACATTGAC-CGGCCGTTCCTTGAGAAGCTGATG------GAATTTTGTGATAATGACAA 2845
Qy    528 gArgPheGlyProAspArgPhePheGluIleLeu------------------------- 539
Db   2846 CAGCTTCGGCAAAGATCGGGCTCGCCTCATCCTCGAATACGTCGCCGATGAGTATGCTGG 2905
Qy    539 ------------------------------------------------------------ 539
Db   2906 AAAGCGTATCTTCGACCCGATGGAACTCTTCAAGGATCATGCAGCCTTAGCTTATTTCCC 2965
Qy    540 -----------------------------------------Ile-ProSerProThrS 545
Db   2966 AACTAGTTTCATGATACCCAACCATTGCGCTTGGGTTCGTAGGGTCACCATCACCCCTAC 3025
Qy    545 erThrSerProSerValProProValValSerLysGlnProGlyAlaValGluGluValI 565
Db   3026 TCGTATCTACTTCAGTACCCCCTGC-GTCGAACCCACG--------------------- 3062
Qy    565 leGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPhePheA 585
Db   3063 -------------------------AACCGCGTCATCCGCCAGTGGAAA------CATG 3090
Qy    585 laLysAsp-----------AlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgA 601
Db   3091 CACAAGACTACTTTATCCGCATTCAATTCACCGATGAAGTTCTTGAAGGACGTATCAAGA 3150
Qy    601 laGluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIleT 621
Db   3151 GCGGCGAAGCGGAACTTCCCCTATTCTTGCGTGTGTACAGGGTATTGGAGAAAGGGGTTG 3210
Qy    621 hrPheArgProAspValPheLysThrArgSerValValProAlaGluGluProValGluG 641
Db   3211 CCATGGGTCCG------TGGCATTGGAAGTTCCTTGCATTTGGCAACTCTCAGATTCGAG 3264
Qy    641 lnArgThrGluPheLysValSerGln----------------------MetLeuAspT 653
Db   3265 AAGCCGGAGCCTTCATGTTCTGTGAACAGAGCAATTTGACAGGCGATATGATGCGAGCCT 3324
Qy    653 rpLeuLeuGlnLeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleG 673
Db   3325 GGATGGGCCGCTTCAGT-----------------CACATCAAGGTCATTGCCAAGTATG 3366
Qy    673 lnLeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisH 693
Db   3367 CGGCTAGACTTGGGCAGTGTTTCTCAACTACCCGTTTGGTGCCTGGCATCCCGGCACCGA 3426
Qy    693 isLysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyA 713
Db   3427 GGATCGTCACCATCCCTGATGTCGAAAAGGACGGATTTTGCTTCACGGATGGAGTAGGCA 3486
Qy    713 rgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspValProS 733
Db   3487 AGATTTCCCCTTTGCTTGCAAAGATCGTAGCGCATGACTGGAGCATTGACCCGCCGCCGT 3546
Qy    733 erAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAspAspT 753
Db   3547 CAGCTTACCAGTTCCGCATGGAGGTTGCAAGGGAGTCTTGGTAACTTGGCCTGATGTGA 3606
Qy    753 hrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAspPheV 773
Db   3607 AGGGTATGGAG------GTTCATATTCGGAAGTCGCAGGAAAAGTTCGTTGCCGAATTC- 3659
Qy    773 alAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaG 793
Db   3660 -------------AATGGCCTTGAGGTTGTCAGA---TGCTCGCAGTTCTCAACCGCCA 3702
Qy    793 lyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLysMetA 813
Db   3703 CTCTCAACCGGCAAATCATTGCCGTTCTCTCGTCT----------------------- 3737
Qy    813 rgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysH 833
Db   3738 --CTTGGTGTTCCCGATCAAGTCTTTGTAGACATGATGGAGCAG---------CAGCTTT 3786
Qy    833 isAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgA 853
Db   3787 CTGACTTCAATGCGGCGATGGAGGACAAACAA----------------------AAGG 3822
Qy    853 laThrArgVal-----------------SerHisGlyArgValProPheLeuAlaGlyL 867
Db   3823 CGACGGCGATTCTGAAAACGTTCATTGATGAAAATCATATGACACCTATCATTGCG---- 3878
Qy    867 euProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysL 887
Db   3879 -----------------GAGATGCTTGCCTAC-----------GGCTTTATGGGGTCTC 3909
Qy    887 ysGlnLysTyrLeuGlnAspIle---------------AlaTrpAspLeuGlnLysArgL 902
Db   3910 AGGAGCCTTTTGTTCGCACCCTACTTCAGCTGTGGCGATCATGGTCGATCAAA------- 3962
Qy    902 ysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrM 922
Db   3963 --------ACGTTGAAGGAGAAAGCGCGACTCAATGTCGAAAAGAGTGCGTTCGTGCTCG 4014
Qy    922 etIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerL 942
Db   4015 GCTGTGTTGATGAGACTGGCACTCTGAAAGGG---------CACATGAAGGTCATTGAAG 4065
Qy    942 ysPheArgAspGluGlu---------------GluSerPheThrLeuLeuSerAspCysA 957
Db   4066 ACTGGAAAGACGTGTCCAGTGAGAAGCTTCCACAGATCTTCCTACAGATTCCCGAT---G 4122
Qy    957 spVal------------------------------LeuValAlaArgSerProAlaH 966
Db   4123 ATGTCAATGGCGGGTACAGGGTTATTACTGGAACCTGCGTCGTTGGCAGAAACCCGTCTC 4182
Qy    966 isPheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuL 986
Db   4183 TCCATCCTGGTGATATCCGTGTGGTGGAGGCAGTTGATGTCCCTGCTTTGCGGCATCTAA 4242
Qy    986 ysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyG 1006
Db   4243 GAGACGTGGTGGTGTTCCCCTTGACTGGGGACAGAGACGTGCCCAGCATGTGCTCTGGTG 4302
Qy   1006 lyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheV 1026
Db   4303 GAGACCTCGATGGTGATGACTTCTTTGTTATTTGGGATCCTTTGCTTATC---------- 4352
Qy   1026 alAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrT 1046
Db   4352 ------------------------------------------------------------ 4352
Qy   1046 hrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrT 1066
Db   4353 ----------------------------------------CCGAAAGAACGCTCCCATC 4371
Qy   1066 yrAspMetIleGlnLysSerPhe---------HisPheAlaLeuGlnProAsnPheLeuG 1083
Db   4372 CCCCAATGATCAGTGAGCCCATTGCGGGCAAGGAGTTGGCAACCGAACCGACCGTAAACA 4431
Qy   1083 lyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysP 1103
Db   4432 ACCTGATTACCTTCTTTGTGTTGTATATGAAGTAC---------------AACAATCTGC 4476
Qy   1103 roAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleV 1123
Db   4477 CGTTGATCGCTCATGCTCATCTGGCG------ACAGCAGATGCGGAGGTGGAAGGAGTGA 4530
Qy   1123 alPheAsnGluAlaSer-----------------TrpAlaGlnLeuArgArgGluLeuL 1137
Db   4531 AGAGCCCGAAATGTAAGTGAACTGTATCATGTTTTTGG---------AAGAAATACCTAT 4581
Qy   1137 euGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArgG 1157
Db   4582 TG--------------------------ACCATTCAACCTGACAGGCCTCGAGCTCGCT 4614
```

```
Qy       1157 lyGluProThrHis-IleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAsp 1176
                          |||      :::||||||:::|||        :::|||
Db       4615 TCACTTCACTCCATGGCTGTTGACTACGTGAAGACTGGTGTTGCCGCC------------ 4662

Qy       1177 LysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAla 1196
                                                          |||     :::
Db       4663 ------------------------GAGTTTCCGAGACGGCTTGACCCCAAGACCTGGCCT 4698

Qy       1197 HisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGluIleSerAspLys 1216
              ||||||    :::    :::        :::|||::::::
Db       4699 CATTTCATGGAGAAAAACAGACACACTTACCACTCAGTA--------------------- 4737

Qy       1217 SerArgSerSerAlaLeu------LeuPheThrThrLeuLysAsnArgIleGlyGluVal 1234
                       :::||||||        |||:::      :::|||            ::::::
Db       4738 ---------ACAGCACTTGGAAAGCTCTACGATATGGTTAAACGGGAAACATTCGACATG 4788

Qy       1235 GluLysGluTyr---------------GlyArgLeuValLysAsn---------LysGlu 1246
              ::::::   |||                     |||::::::|||:::          :::
Db       4789 AAAGAGAACTACCAGCTGCCGTTTGATAACCGCATCCTCAAGCACACAAAGTGCCGCGCT 4848

Qy       1247 MetArgAspSer-----------------------LysAspProTyrProValArgVal 1258
              :::||||||                               |||      |||         :::
Db       4849 CTTAGGGATGGGACGCTGGCCAAAGCCCGTCGCATCAAATCTCAGTACGACACCGCTATG 4908

Qy       1259 AsnGlnVal---------------------------TyrGluLysTrpCysAlaIleThr 1269
                 :::|||                            :::|||   |||   |||
Db       4909 CGCCGCGTGATGTGCCAGCTCGAAATTGCGACCGAGTTCGAAGTTTGGACCGCCTTT--- 4965

Qy       1270 ProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysValIleArgLeuLeuGlu 1289
                       |||   |||                  |||   |||
Db       4966 ------GTCATGTCCAAACCCCGCGTGGGCTCAGACTACAAA------------------ 5001

Qy       1290 LeuSerPheLeuAlaAspArgGluMetAsnThrTrpAlaLeuLeuArgAlaSerThrAla 1309
                       |||   |||                      :::    |||    ||:::|||
Db       5002 ---------CTCCAAGACAAC------------------GTCGGTCGCGAGTCTTCTGCC 5034

Qy       1310 PheLysLeuTyrTyrHisLysSerProLysPheValTrpGlnMetAlaGlyArgGlnLeu 1329
                 |||   ::::::             |||         :::       |||      |||
Db       5035 CTCAAGCAACACTTCAAGGATCAGTGTAAA------AAGGAGGCTGGGGGTGATCTCCTT 5088

Qy       1330 AlaTyrIleLysAlaGln-MetThrSerArgProGlyGluGlyAla-------------- 1344
              :::::::::   |||    :::    |||         :::    |||
Db       5089 TCCTTCGTATCGGCCATGTATAGAGTCACTTACGAAGAAGTTCGCATTGCCCTTTTCGAA 5148

Qy       1345 -ProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLys 1360
                |||:::      |||       :::    ||||||       |||:::|||
Db       5149 GCCAAGCAGCCGCACGTTAGACCTGACGGCAGGCTCGGAACCAGAAAAA 5197
```

```
RESULT 8
DDI314910
LOCUS       DDI314910            5992 bp    DNA     linear   INV 14-JUN-2001
DEFINITION  Dictyostelium discoideum partial RrpB gene for putative RNA
            dependent RNA polymerase, exons 1-2.
ACCESSION   AJ314910
VERSION     AJ314910.1  GI:14475570
KEYWORDS    RNA dependent RNA polymerase; RrpB gene.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1
  AUTHORS   Martens,H., Novotny,J., Oberstrass,J., Steck,T.L., Postlethwait,P.
            and Nellen,W.
  TITLE     RNAi in Dictyostelium: developmental regulation and the role of
            RdRPs and dsRNase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5992)
  AUTHORS   Martens,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2001) Martens H., Genetics Department, Kassel
            University, Heinrich-Plett Strasse 40, 34132, GERMANY
FEATURES             Location/Qualifiers
     source          1. .5992
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
     gene            1. .5992
                     /gene="RrpB"
     mRNA            join(<1. .970,1067. .5992)
                     /gene="RrpB"
     CDS             join(<1. .970,1067. .5992)
                     /gene="RrpB"
                     /codon_start=2
                     /product="putative RNA dependent RNA polymerase"
                     /protein_id="CAC41975.1"
                     /db_xref="GI:14475571"
                     /db_xref="GOA:Q95ZG6"
                     /db_xref="SPTREMBL:Q95ZG6"
                     /translation="RLTIEIVQRSIKDMEERMLSRVFKPHSLLTNSTQSELQPELVSF
                     KTSGQESLIETSICEFLVSNSKELQIFLNYSDVDLDGVRGNPSFLKALERLDKYSKQL
                     ESIYIEYTKVLIKLYELLSVLSTEGPKQVLTGLDLLMQSTNQENELYNKLLELRYYIS
                     GQYEDRESNFEKGSTRYRKLISSLEYAISDASDGELQKDLRILVFVETRSGASNLTSM
                     LKKEPFQEYLHTKRLVGHNGDDGMDSEKQQSIIRKFRDGKCRLIVTTNVLEEGIDVQD
                     CNIVICYDGILSLKSLIQRRGRARSKNESKFIIIYNDDKDLKILNILSSENLLNNSIN
                     QILSNRERQAITKNQFLNWKKLNINNEYFINNQLDGGPLDKDIKCVILSIYHIEIDIE
                     TFRELLLFQLPDIVNQVDQVSIRMEGDDEYLSSSIIVILNTCIDKEQSIYRLKSLLNI
                     NGLKYWMIDRSSFHFEDDNIDNDNNNSYDRENEQDNFNYFINEYTHDLSQFKNKNYSI
                     GSFEYGNMSSPTKYHGIGQLDDSFDLFYGNREMFFTKKTRFQRYRLNLKLTDIESTFL
                     LVNDKYQNIVHILLIVKRPLIFEEEKETTHRVFWNKSKPKKDDFGQAFVYRLTLSTID
                     NEIKEVIKPLIANGFRMYESSLELNCFIQGGINKNNNCMIIDLNLQYIKRNLEIHYLL
                     QCLQSQSSYGRVLNQDFIVEINNLLAFGKQEEAEFIITSLISKRTKFINYKEFIGDCL
                     KNFLPDPYYAKKKDLSTKSGEFTMIRDVFITPSRTIFIQPTLQRSCRAVRKFGSSNFI
                     MVKIVNESLEPLQGVNNKHPLHENRIKPILEGGILVGGKIYSYAGNSNSQLREYSSWF
                     VSNQIGTHTVKIWSGIEHVDNVRKFFRCIGLMFSTTIPTVTLPQNRIYRIQDITRNTH
                     VFTEGCGEIGPELAKHLNENYNFRPSTCAYQVRIGGNKGMLVVNNQAPDPSGIYIRPS
                     MVKFNPIVCGDEHRTLEICSVSTTSRCKLNRQVISLLSTLGTQDNVFFALQDHYLNQV
                     AQIVNDTNASKQAIVEFFPDITEGELYQDPYIRRILISLYKLKMERIQQKCHIEIKDS
                     RMLLGVCDPTNSLPPNTVFVQLEEEDEDDDDGRKYEKVIEGLVMVIKNPCTHPGDVRY
                     LKAVDNLRLRHLRNVLVFSTKGDVPNFKEISGSDLDGDRYFFCYDKSLIGNRSKSETA
                     YLGDETVSNNDKKANVFNDPFALSSMYSTNVERQELGKMYNSHLAISDLFGANHEFSI
                     QISKECFKEIDYPKTGIHGTIPKEANVWLKTKGYPHYMQRENSTRVYYQSKTIMGKMY
                     DQIDQLVYIGDFLPNISLDKSNLVDDYEIYLNSAKILYSQYKLQVHSLLRHYSAESEE
                     SIMIGFLDQGFISDKVSKDIKGEMKNDYIKIQQTFENEFLKEFGEQHKENCLLIHRVN
                     IEKKVSAWYHVAYSDLKDDRALGFYWIARSFKQMTLDQESQKKKNILVNSILDHMNKK
                     KDTLFPLYESRLMIVKELTLFLNMLNDTVFKESTLEIIGSTATMLFDENSNLNLYLKL
                     ESQNDHMNLAEKNEILNQLKDEIKTSSFISAEACSIIKIQGDIPLFKFKIDKIQCLIG
                     LETSKWHETLVQQQYITKNPSILSILFSIIDWSINSLFVDSETGEIINNSKVYSSLNL
                     QKVIVENSMIPSIFNNKNHNNNNNNNNNKHSVKNLTKSNLFSLLIDFCIINKYIDKC
                     NPNELTNTTSNLTIGNWVKLIEIASRYFSVDNNNNNNNKGHKRKKNKNLDEYNTNNLG
                     DTILSFFRFYSDSFGRSIYGNEPFIIKLFSKDANDIENQILNPSTESHKTALLFYQES
                     FRHAFHSLSTFVSVEEFLSNCLKTTKKIMPIKKNQINLLQERSVNEEIVKNVQLCGCS
                     IKIDTNVKNPYILIDGKPQQIRLAQNAIRDLLSIKTENYNLAITKRNNKRKNFTLGDF
                     AKNKIK"
     exon            <1. .970
                     /gene="RrpB"
                     /number=1
     intron          971. .1066
                     /gene="RrpB"
                     /number=1
     exon            1067. .5992
                     /gene="RrpB"
                     /number=2
ORIGIN

Alignment Scores:
Pred. No.:              0.000347       Length:         5992
Score:                  298.00         Matches:        173
Percent Similarity:     35.68%         Conservative:   136
Best Local Similarity:  19.98%         Mismatches:     291
Query Match:            4.03%          Indels:         266
DB:                     3              Gaps:           39

US-09-913-878A-2 (1-1402) x DDI314910 (1-5992)

Qy          509 ProLeuTyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrArg 528
```

```
Db    2411 CCAAGTAGAACTATTTCATTCAACCAACACTA---CAAAGATCATGTAGAGCTGTTCGT 2467
Qy     529 ArgPheGlyProAspArgPhePheGluIleLeuIleProSerProThrSerThrSerPro 548
Db    2468 AAATTTGGTTCATCCAATTTCATTATGGTTAAAATT---------------GTAAATGAA 2512
Qy     549 SerValProProValValSerLysGlnProGlyAlaValGluGluValIleGlnTrpLeu 568
Db    2513 TCATTGGAACCACTT--------------------------------------------- 2527
Qy     569 ThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPhePheAlaLysAspAla 588
Db    2528 ---------------------------------------------------CAAGGTGTA 2536
Qy     589 GlyTyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysProIleIle 608
Db    2537 AATAATAAACATCCACTTCATGAGAATAGAATT---------------AAACCAATACTT 2581
Qy     609 LysGluArgValHisPhePheAlaGluThrGlyIleThrPheArgProAspValPheLys 628
Db    2582 ------------------------GAGGGTGGTATATTGGTTGGTGGTAAAATCTATTCG 2617
Qy     629 ThrArgSerValValProAlaGluGluProValGluGlnArgThrGluPheLysValSer 648
Db    2618 TATGCTGGT-----------------------------AATAGTAATAGTCAGTTAAGA 2647
Qy     649 GlnMetLeuAspTrpLeuLeu------GlnLeuAspAsnAsnThr---------TrpGln 663
Db    2648 GAATATTCATCTTGGTTCGTTTCTAATCAAATTGGTACACATACGGTTAAAATTTGGAGT 2707
Qy     664 ------ProHisLeu------------LysLeuPheSerArgIleGlnLeuGlyLeuSer 677
Db    2708 GGTATAGAGCATGTTGATAATGTTAGAAAGTTTTTCAGATGTATTGGTTTAATGTTTTCA 2767
Qy     678 LysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAspLeu 697
Db    2768 ACAACGATTCCAACGGTTACATTACCACAGAATCGTATC---TATAGAATCCAAGATATC 2824
Qy     698 LeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSer 717
Db    2825 ---------ACCAGAAATACCCATGTATTCACTGAGGGATGTGGTGAAATTGGTCCAGAA 2875
Qy     718 ValAlaLysArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGly 737
Db    2876 CTTGCTAAACATTTAAATGAAAATTATAATTTCAGACCATCAACATGTGCATATCAAGTT 2935
Qy     738 ArgPheGlySerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAsp 757
Db    2936 AGAATAGGTGGAAATAAAGGTATGTTGGTGGTAAAC---AATCAAGCACCCGATCCATCT 2992
Qy     758 TrpIleGluThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGln 777
Db    2993 GGTATTTATATTAGACCAAGTATGGTGAAATTCAATCCAATCGTTTGTGGTGATGAACAT 3052
Qy     778 ArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGln 797
Db    3053 AGAACATTAGAGATTTGCAGTGTT---TCAACAACTTCAAGATGTAAATTAAATAGACAA 3109
Qy     798 LeuLeuProValLeuGluAsp---ArgAlaArgAspLysValLysMetArgGlnAlaIle 816
Db    3110 GTAATTAGTTTACTCTCAACATTGGGTACACAAGATAATGTT------------------ 3151
Qy     817 GlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsn 836
Db    3152 ------------------------------TTCTTTGCATTACAAGATCATTATCTAAAT 3181
Qy     837 ArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgVal 856
Db    3182 ---------CAAGTGGCTCAAATCGTAAATGATACCAATGCTTCCAAACAAGCAATCGTT 3232
Qy     857 SerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsn 876

Db    3233 GAATTC-----------------------TTTCCAGATATTACAGAGGGTGAACTT--- 3265
Qy     877 PheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrp 896
Db    3266 --------------------------TATCAAGATCCTTACATTCGTAGAATTTTAATT 3298
Qy     897 AspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArg 916
Db    3299 AGTCTTTATAAATTAAAAATGGAGAGAATTCAACAAAAATGTCATATAGAGATTAAAGAT 3358
Qy     917 SerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHis 936
Db    3359 AGTAGAATGTTGTTGGGTGTTTGTGATCCAACAAATTCATTACCACCAAATACTGTATTT 3418
Qy     937 ValGlyPheSerSerLysPheArgAspGluGluGlu------SerPheThrLeuLeuSer 954
Db    3419 GTTCAATTGGAAGAGGAAGATGAAGATGATGATGATGGTCGTAAATATGAAAAAGTAATT 3478
Qy     955 AspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgVal 974
Db    3479 GAAGGATTGGTAATGGTAATTAAAAATCCTTGTACTCATCCTGGTGATGTAAGATATTTA 3538
Qy     975 ArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLys 994
Db    3539 AAGGCTGTTGATAATTTAAGATTAAGACATTTAAGAAATGTATTAGTTTTCTCAACAAAA 3598
Qy     995 GlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrp 1014
Db    3599 GGTGACGTACCAAACTTTAAAGAAATTTCAGGTTCAGATTTAGATGGTGATCGTTATTTC 3658
Qy    1015 ValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGluPro 1034
Db    3659 TTTTGTTACGATAAATCATTGATT------------------------------------ 3682
Qy    1035 AspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHis 1054
Db    3682 ------------------------------------------------------------ 3682
Qy    1055 GlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHis 1074
Db    3683 ------GGTAATCGTTCAAAATCAGAGACAGCCTAT------------------------ 3712
Qy    1075 PheAlaLeuGlnProAsnPheLeuGly---------MetCysThrAsnTyrLysGluArg 1091
Db    3713 ---------------------TTAGGTGACGAGACCGTTTCAAACAATGATAAAAAGGCC 3751
Qy    1092 LeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuVal 1111
Db    3752 ------------------AATGTTTTCAATGATCCA---TTTGCTTTATCCTCAATGTAC 3790
Qy    1112 GlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGln 1131
Db    3791 TCTACAAATGTTGAACGTCAAGAACTTGGTAAAATGTATAAT------------TCACAT 3838
Qy    1132 LeuArgArgGluLeuLeuGlyGlyAla---------------LeuSerLeuProAspPro 1146
Db    3839 TTGGCAATCTCTGATCTTTTCGGTGCAAATCACGAATTTTCAATTCAAATCTCAAAAGAA 3898
Qy    1147 MetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeu 1166
Db    3899 TGTTTTAAAGAA------------------------------------ATTGATTATCCA 3922
Qy    1167 LysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMet 1186
Db    3923 AAAACTGGTATT---CATGGCACTATTCCAAAAGAAGCAAATGTTTGG------------ 3967
Qy    1187 LysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSer--- 1205
Db    3968 ---------TTAAAAACTAAAGGTTATCCACATTATATGCAAAGAGAGAATTCAACCAGA 4018
Qy    1206 ---TyrTyr--------------------------------------------------- 1207
Db    4019 GTTTATTATCAATCGAAAACCATTATGGGTAAAATGTATGATCAAATTGATCAGTTGGT 4078
```

```
Qy   1208 ---------ThrPhePheLysGluIleSer---AspLysSer----------------- 1217
Db   4079 TATATTGGTGATTTCTTACCAAATATTAGTCTTGATAAAAGCAATTTGGTTGATGACTAT 4138
Qy   1218 ------------ArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGlu 1233
Db   4139 GAAATCTATTTAAATAGTGCTAAAATCTTATATAGTCAGTATAAACTTCAAGTTCATTCA 4198
Qy   1234 ValGluLysGluTyrGlyArg--------------------------------------- 1240
Db   4199 TTGTTAAGACATTATTCAGCGGAGTCTGAAGAATCTATAATGATTGGTTTTCTTGATCAA 4258
Qy   1241 ---LeuValLysAsnLysGluMetArgAsp-----------SerLysAspProTyrPro 1255
Db   4259 GGTTTTATTAGTGACAAAGTTTCAAAAGATATCAAAGGTGAAATGAAAAATGATTAC--- 4315
Qy   1256 ValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrProGluAlaMetAspLys 1275
Db   4316 ATTAAAATTCAACAAACTTTTGAAAAC-----------------GAGTTTTTAAAGGAA 4357
Qy   1276 SerGlyAlaAsnTyrAspSerLysValIleArgLeuLeuGluLeuSerPheLeuAlaAsp 1295
Db   4358 TTTGGTGAACAACATAAAGAGAATTGTTTATTGATTCATAGAGTAAACATT------GAA 4411
Qy   1296 ArgGluMetAsnThrTrp 1301
Db   4412 AAGAAGGTATCAGCATGG 4429
```

```
RESULT 9
AR145904
LOCUS       AR145904             3731 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 1 from patent US 6218142.
ACCESSION   AR145904
VERSION     AR145904.1  GI:15109093
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3731)
  AUTHORS   Wassenegger,M., Riedel,L., Schiebel,W. and Sanger,H.L.
  TITLE     Nucleic acid molecules encoding polypeptides having the enzymatic
            activity of an RNA-directed RNA polymerase (RDRP)
  JOURNAL   Patent: US 6218142-A 1 17-APR-2001;
FEATURES             Location/Qualifiers
     source          1..3731
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.000197      Length:        3731
Score:                  297.50        Matches:       189
Percent Similarity:     35.78%        Conservative:  137
Best Local Similarity:  20.75%        Mismatches:    339
Query Match:            4.02%         Indels:        247
DB:                     6             Gaps:          35

US-09-913-878A-2 (1-1402) x AR145904 (1-3731)
```

```
Qy    562 GluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArg 581
Db   1349 GAAGACATAGATAACTTTCTTCGTGTT---------TCTTTTGTTGATGAGGAGTGGGAG 1399
Qy    582 AlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAla 601
Db   1400 AAACTGTATTCTACAGACTTA--------------------------------------- 1420
Qy    602 GluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIleThr 621
Db   1421 ---TTACCAAAA-----------------------------GCAAGTACTGGAAGTGGT 1447
Qy    622 PheArgProAspValPheLys-----------------ThrArgSerValValProAla 635
Db   1448 GTCAGGACAAACATCTATGAGAGGATCTTATCAACTCTGCGGAAAGGCTTTGTAATTGGT 1507
Qy    636 GluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeu 655
Db   1508 GATAAAAAATTTGAA---TTTCTTGCATTTTCATCGAGCCAGTTG--------------- 1549
Qy    656 GlnLeuAspAsnAsnThrTrp------------GlnProHisLeuLysLeu--------- 668
Db   1550 ---CGGGATAATTCAGTGTGGATGTTTGCATCAAGACCTGGCCTTACTGCAAATGATATA 1606
Qy    669 -----------------PheSerArgIle-----------------------GlnLeu 674
Db   1607 AGAGCTTGGATGGGTGATTTTTCGCAGATCAAGAATGTCGCAAAATATGCTGCCAGACTT 1666
Qy    675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLys 694
Db   1667 GGTCAATCTTTTGGTTCCTCCAGAGAGACTTTG---------AGTGTTCTTAGGCATGAG 1717
Qy    695 ThrAspLeuLeu---------SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
Db   1718 ATTGAAGTTATTCCCGATGTAAAGGTTCATGGAACCAGCTATGTCTTTTCTGATGGAATT 1777
Qy    712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal 731
Db   1778 GGTAAAATATCTGGTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCTTCAATATACC 1837
Qy    732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAsp 751
Db   1838 CCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGGTGTTGATCCGGAT 1897
Qy    752 AspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAsp 771
Db   1898 TCATCAATGAAG---------TTGTCTTTGAGAAAGAGCATGTCGAAATATGAATCAGAC 1948
Qy    772 PheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791
Db   1949 AACATA---------------AAGTTAGATGTCCTTGGA---TGGAGCAAATATCAGCCT 1990
Qy    792 AlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLys 811
Db   1991 TGTTATCTTAATCGTCAACTGATTACGCTCTTGTCT------------------------ 2026
Qy    812 MetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGln 831
Db   2027 ---ACACTTGGAGTGAAAGATGAAGTTCTCGAACAGAAGCAAAAGGAAGCTGTAGATCAG 2083
Qy    832 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 851
Db   2084 CTTGATGCTATCTTGCATGATTCTTTGAAGGCACAGGAGGCTTTGGAATTGATGTCT--- 2140
Qy    852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
Db   2141 ------------------------------------------------CCTGGAGAGAAC 2152
Qy    872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeu 891
Db   2153 ACTAATATTCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCCTTTCTT 2212
Qy    892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
Db   2213 TCAATGATGTTGCAAACCTTCCGCGCATCCAAGTTGCTCGATTTGCGGACTAGATCAAGA 2272
Qy    912 IleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGlu 931
Db   2273 ATATTTATTCCAAATGGAAGAACAATGATGGGATGTTTGGATGAATCCAGAACCTTGGAA 2332
Qy    932 GluAsnGluValHisValGlyPheSer------------SerLysPheArgAspGluGlu 947
Db   2333 TATGGTCAGGTGTTTGTTCAGTTTACTGGTGCTGGACATGGAGAGTTTTCTGACGATTTA 2392
Qy    948 GluSerPhe-----------------------ThrLeuLeuSerAspCysAspValLeu 959
```

```
Db     2393 CATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAATGTGGTT 2452

Qy      960 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 979
Db     2453 GTTGCAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTTAAAGGCTGTAAATGTT 2512

Qy      980 ProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeu 999
Db     2513 CGAGCGCTGCACCACATGGTAGATTGTGTTGTATTCCCTCAGAAAGGAAAAAGACCTCAT 2572

Qy     1000 AlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspPro 1019
Db     2573 CCGAATGAATGTTCTGGGAGTGATTTGGATGGGGATATCTACTTTGTTTGCTGGGATCAA 2632

Qy     1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
Db     2633 GACATGATCCCGCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCCAGCATACAG 2692

Qy     1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
Db     2693 TTGGACCATGAT------------------------------------------------ 2704

Qy     1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 1079
Db     2705 ------------GTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATTGTGAAT 2752

Qy     1080 AsnPheLeuGlyMetCysThrAsn------------TyrLysGluArgLeuCysTyrIle 1095
Db     2753 GACAGTTTGGGAATCATAGCAAATGCCCATGTCGTATTTGCAGACAGA---------GAA 2803

Qy     1096 AsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVal 1115
Db     2804 CCTGATATGGCCATGAGTGATCCATGCAAAAAACTTGCTGAGCTCTTTTCAATTGCAGTG 2863

Qy     1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 1135
Db     2864 GACTTTCCAAAGACTGGTGTTCCCGCTGAAATACCATCT------CAGTTGCGC------ 2911

Qy     1136 LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTrpLeuGl 1155
Db     2912 ------------------------------CCTAAAGAATACCCAGACTTCATGGATAAG 2941

Qy     1155 yArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIl 1175
Db     2942 CCGG------------------------------------------ACAAGACCAGCTAT 2959

Qy     1175 eAspLysGlu------LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGl 1193
Db     2960 ATCTCAGAAAGAGTTATTGGAAAGCTTTTC-AGGAAAGTGAAG----------------- 3001

Qy     1193 uAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGluIl 1213
Db     3002 -GACAAAGCA------------CCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGT 3048

Qy     1213 eSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGl 1233
Db     3049 TGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACATTGACGA 3108

Qy     1233 uVal---------GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSe 1250
Db     3109 AGCTTTTGACTACAAAACTGAATAT---------------GACAACAAGCTGGGTAATTT 3153

Qy     1250 rLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrPr 1270
Db     3154 AATGGACTACTATGGCATAAAAACAGAG------------------GCTGAAATACTTAG 3195

Qy     1270 oGluAlaMetAspLysSerGlyAlaAsnTyrAsp------------------------Se 1282
Db     3196 TGGTGGCATTATGAAGGCATCAAAAACTTTTGACCGCAGAAAAGATGCTGAGGCCATTAG 3255

Qy     1282 rLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp-- 1301
Db     3256 TGTTGCTGTGAGGGCCTTGAGG------------------AAGGAGGCAAGAGCCTGGTT 3297

Qy     1302 ------------------------AlaLeuLeuArgAlaSerThrAlaPheLysLeuTy 1313
Db     3298 CAAGAGGCGTAATGATATAGATGACATGTTACCAAAGGCTTCGGCTTGGTACCACGTTAC 3357

Qy     1313 rTyrHisLysSer------------------------------------------ProLy 1319
Db     3358 ATATCATCCTACATATTGGGGTTGCTACAATCAGGGGTTGAAAAGAGCTCATTTCATTAG 3417

Qy     1319 sPheValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerAr 1339
Db     3418 CTTTCCCTGGTGTGTTTATGACCAGCTAATCCAGATTAAGAAGGACAAAGCACGTAACAG 3477

Qy     1339 gProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAs 1359
Db     3478 GCCAGTTCTCAACTTGTCATCTCTCAGGGCTCAACTGAGTCACAGATTAGTG-------- 3529

Qy     1359 pLysLysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAs 1379
Db     3530 ------------------------------TTGAAATGAGATTCCAGTCGAGCGTTAAG 3558

Qy     1379 pProGluValTyrGluValLeuGlyAspAsp 1389
Db     3559 CTGATATATATATAATGTAATAGGGTGTGAT 3589
```

```
RESULT 10
LERDRP
LOCUS       LERDRP                  3731 bp    mRNA    linear   PLN 18-DEC-1998
DEFINITION  L.esculentum mRNA for RNA-directed RNA polymerase.
ACCESSION   Y10403
VERSION     Y10403.1  GI:4038591
KEYWORDS    RdRP gene; RNA-directed RNA polymerase.
SOURCE      Lycopersicon esculentum (tomato)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1
  AUTHORS   Schiebel,W., Pelissier,T., Riedel,L., Thalmeir,S., Schiebel,R.,
            Kempe,D., Lottspeich,F., Sanger,H.L. and Wassenegger,M.
  TITLE     Isolation of an RNA-directed RNA polymerase-specific cDNA clone
            from tomato
  JOURNAL   Plant Cell 10 (12), 2087-2101 (1998)
  MEDLINE   99055198
   PUBMED   9836747
REFERENCE   2  (bases 1 to 3731)
  AUTHORS   Wassenegger,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-1997) M. Wassenegger, Max-Planck-inst. fuer
            Biochemie, Viroidforschung, Am Klopferspitz 18a,
            Planegg-Martinsried, 82152, FRG
FEATURES             Location/Qualifiers
     source          1..3731
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /strain="Rutgers"
                     /db_xref="taxon:4081"
     gene            1..3731
                     /gene="RdRP"
     CDS             194..3538
                     /gene="RdRP"
                     /EC_number="2.7.7.48"
                     /codon_start=1
                     /product="RNA-directed RNA polymerase"
                     /protein_id="CAA71421.1"
                     /db_xref="GI:4038592"
                     /db_xref="GOA:Q9ZR58"
                     /db_xref="SPTREMBL:Q9ZR58"
                     /translation="MGKTIQVFGFPYLLSAEVVKSFLEKYTGYGTVCALEVKQSKGGS
                     RAFAKVQFADNISADKIITLANNRLYFGSSYLKAWEMKTDIVQLRAYVDQMDGITLNF
                     GCQISDDKFAVLGSTEVSIQFGIGLKKFFFFLSSGSADYKLQLSYENIWQVVLHRPYG
                     QNAQFLLIQLFGAPRIYKRLENSCYSFFKETPDDQWVRTTDFPPSWIGLSSSLCLQFR
                     RGVRLPNFEESFFHYAERENNITLQTGFTFFVSQKSALVPNVQPPEGISIPYKILFKI
```

```
                    SSLVQHGCIPGPALNVYFFRLVDPRRRNVACIEHALEKLYYIKECCYDPVRWLTEQYD
                    GYLKGRQPPKSPSITLDDGLVYVRRVLVTPCKVYFCGPEVNVSNRVLRNYSEDIDNFL
                    RVSFVDEEWEKLYSTDLLPKASTGSGVRTNIYERILSTLRKGFVIGDKKFEFLAFSSS
                    QLRDNSVWMFASRPGLTANDIRAWMGDFSQIKNVAKYAARLGQSFGSSRETLSVLRHE
                    IEVIPDVKVHGTSYVFSDGIGKISGDFAHRVASKCGLQYTPSAFQIRYGGYKGVVGVD
                    PDSSMKLSLRKSMSKYESDNIKLDVLGWSKYQPCYLNRQLITLLSTLGVKDEVLEQKQ
                    KEAVDQLDAILHDSLKAQEALELMSPGENTNILKAMLNCGYKPDAEPFLSMMLQTFRA
                    SKLLDLRTRSRIFIPNGRTMMGCLDESRTLEYGQVFVQFTGAGHGEFSDDLHPFNNSR
                    STNSNFILKGNVVVAKNPCLHPGDIRVLKAVNVRALHHMVDCVVFPQKGKRPHPNECS
                    GSDLDGDIYFVCWDQDMIPPRQVQPMEYPPAPSIQLDHDVTIEEVEEYFTNYIVNDSL
                    GIIANAHVVFADREPDMAMSDPCKKLAELFSIAVDFPKTGVPAEIPSQLRPKEYPDFM
                    DKPDKTSYISERVIGKLFRKVKDKAPQASSIATFTRDVARRSYDADMEVDGFEDYIDE
                    AFDYKTEYDNKLGNLMDYYGIKTEAEILSGGIMKASKTFDRRKDAEAISVAVRALRKE
                    ARAWFKRRNDIDDMLPKASAWYHVTYHPTYWGCYNQGLKRAHFISFPWCVYDQLIQIK
                    KDKARNRPVLNLSSLRAQLSHRLVLK"
ORIGIN

Alignment Scores:
Pred. No.:               0.000197        Length:         3731
Score:                   297.50          Matches:        189
Percent Similarity:      35.78%          Conservative:   137
Best Local Similarity:   20.75%          Mismatches:     339
Query Match:             4.02%           Indels:         247
DB:                      8               Gaps:           35

US-09-913-878A-2 (1-1402) x LERDRP (1-3731)

Qy      562 GluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArg 581
            |||:::::        :::|||    :::        |||   |||     :::|||
Db     1349 GAAGACATAGATAACTTTCTTCGTGTT---------TCTTTTGTTGATGAGGAGTGGGAG 1399

Qy      582 AlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAla 601
                  ::::::   |||
Db     1400 AAACTGTATTCTACAGACTTA--------------------------------------- 1420

Qy      602 GluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIleThr 621
                 ||||||                                  |||   ||||||
Db     1421 ---TTACCAAAA-----------------------------GCAAGTACTGGAAGTGGT 1447

Qy      622 PheArgProAspValPheLys------------------ThrArgSerValValProAla 635
             |||   :::::::::::                      :::      |||
Db     1448 GTCAGGACAAACATCTATGAGAGGATCTTATCAACTCTGCGGAAAGGCTTTGTAATTGGT 1507

Qy      636 GluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeu 655
            ::::::        |||          |||      ||||||:::
Db     1508 GATAAAAAATTTGAA---TTTCTTGCATTTTCATCGAGCCAGTTG--------------- 1549

Qy      656 GlnLeuAspAsnAsnThrTrp------------GlnProHisLeuLysLeu--------- 668
                  ||||||:::   |||            :::|||   |||
Db     1550 ---CGGGATAATTCAGTGTGGATGTTTGCATCAAGACCTGGCCTTACTGCAAATGATATA 1606

Qy      669 ------------------PheSerArgIle-----------------------GlnLeu 674
                              ||||||:::|||                        :::|||
Db     1607 AGAGCTTGGATGGGTGATTTTTCGCAGATCAAGAATGTCGCAAAATATGCTGCCAGACTT 1666

Qy      675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLys 694
            |||   |||         :::      ||||||           :::      |||:::
Db     1667 GGTCAATCTTTTGGTTCCTCCAGAGAGACTTTG---------AGTGTTCTTAGGCATGAG 1717

Qy      695 ThrAspLeuLeu---------SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
               :::::::::                  ||||||    |||   :::||||||:::
Db     1718 ATTGAAGTTATTCCCGATGTAAAGGTTCATGGAACCAGCTATGTCTTTTCTGATGGAATT 1777

Qy      712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal 731
            |||:::::::|||      |||      |||:::             ||||||
Db     1778 GGTAAAATATCTGGTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCTTCAATATACC 1837

Qy      732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAsp 751
            ||||||||   |||   |||:::|||      ||||||:::      :::|||   |||
Db     1838 CCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGGTGTTGATCCGGAT 1897

Qy      752 AspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAsp 771
               :::               :::               |||      |||:::|||   |||
Db     1898 TCATCAATGAAG---------TTGTCTTTGAGAAAGAGCATGTCGAAATATGAATCAGAC 1948

Qy      772 PheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791
               :::                     |||:::|||            |||:::   :::
Db     1949 AACATA---------------AAGTTAGATGTCCTTGGA---TGGAGCAAATATCAGCCT 1990

Qy      792 AlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLys 811
                  ||||||   ||||||:::   :::|||
Db     1991 TGTTATCTTAATCGTCAACTGATTACGCTCTTGTCT------------------------ 2026

Qy      812 MetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGln 831
                  :::   |||      ::::::            |||::::::      :::|||
Db     2027 ---ACACTTGGAGTGAAAGATGAAGTTCTCGAACAGAAGCAAAAGGAAGCTGTAGATCAG 2083

Qy      832 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 851
                  |||:::                           |||      |||      |||
Db     2084 CTTGATGCTATCTTGCATGATTCTTTGAAGGCACAGGAGGCTTTGGAATTGATGTCT--- 2140

Qy      852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
                                                         |||
Db     2141 -------------------------------------------------CCTGGAGAGAAC 2152

Qy      872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeu 891
                     |||   ::::::|||   |||:::   |||      :::   :::|||
Db     2153 ACTAATATTCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCCTTTCTT 2212

Qy      892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
               :::         :::         |||            |||::::::::
Db     2213 TCAATGATGTTGCAAACCTTCCGCGCATCCAAGTTGCTCGATTTGCGGACTAGATCAAGA 2272

Qy      912 IleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGlu 931
            |||   :::               :::            |||               |||||
Db     2273 ATATTTATTCCAAATGGAAGAACAATGATGGGATGTTTGGATGAATCCAGAACCTTGGAA 2332

Qy      932 GluAsnGluValHisValGlyPheSer------------SerLysPheArgAspGluGlu 947
               :::|||   |||   |||:::                  :::|||   |||:::
Db     2333 TATGGTCAGGTGTTTGTTCAGTTTACTGGTGCTGGACATGGAGAGTTTTCTGACGATTTA 2392

Qy      948 GluSerPhe-----------------------ThrLeuLeuSerAspCysAspValLeu 959
                |||                          :::            :::|||:::
Db     2393 CATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAATGTGGTT 2452

Qy      960 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 979
            ||||||:::::::|||         |||   ||||||:::      ::::::||||||
Db     2453 GTTGCAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTTAAAGGCTGTAAATGTT 2512

Qy      980 ProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeu 999
               ||||||   :::   |||   ::::::|||      ||||||         |||
Db     2513 CGAGCGCTGCACCACATGGTAGATTGTGTTGTATTCCCTCAGAAAGGAAAAAGACCTCAT 2572

Qy     1000 AlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspPro 1019
               :::   ||||||   |||   |||||||||:::   :::||||||||||||
Db     2573 CCGAATGAATGTTCTGGGAGTGATTTGGATGGGGATATCTACTTTGTTTGCTGGGATCAA 2632

Qy     1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
            ::::::::                  |||      |||      |||
Db     2633 GACATGATCCCGCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCCAGCATACAG 2692

Qy     1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
            |||      |||
Db     2693 TTGGACCATGAT------------------------------------------------ 2704

Qy     1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 1079
                        |||   :::   :::::::::   |||            :::
Db     2705 ------------GTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATTGTGAAT 2752

Qy     1080 AsnPheLeuGlyMetCysThrAsn------------TyrLysGluArgLeuCysTyrIle 1095
            :::   ||||||:::         |||               :::   :::|||
Db     2753 GACAGTTTGGGAATCATAGCAAATGCCCATGTCGTATTTGCAGACAGA---------GAA 2803

Qy     1096 AsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVal 1115
               :::            :::   |||         |||:::   |||            |||
Db     2804 CCTGATATGGCCATGAGTGATCCATGCAAAAAACTTGCTGAGCTCTTTTCAATTGCAGTG 2863
```

```
Qy      1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 1135
             |||         |||   |||:::               |||         ||||||||
Db      2864 GACTTTCCAAAGACTGGTGTTCCCGCTGAAATACCATCT------CAGTTGCGC------ 2911

Qy      1136 LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTrpLeuGl 1155
                                          |||            ||| ||||||:::
Db      2912 ------------------------------CCTAAAGAATACCCAGACTTCATGGATAAG 2941

Qy      1155 yArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIl 1175
              |||                                          ||||||||||
Db      2942 CCGG-----------------------------------------ACAAGACCAGCTAT 2959

Qy      1175 eAspLysGlu------LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGl 1193
             |  ::::::       ||||||:::|||         :::|||
Db      2960 ATCTCAGAAAGAGTTATTGGAAAGCTTTTC-AGGAAAGTGAAG----------------- 3001

Qy      1193 uAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGluIl 1213
              |||   |||              |||      :::|||     |||||    ::::::::
Db      3002 -GACAAAGCA------------CCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGT 3048

Qy      1213 eSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGl 1233
             ::::   :::|||      :::   :::            ::::::   |||    ||
Db      3049 TGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACATTGACGA 3108

Qy      1233 uVal---------GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSe 1250
             |            :::   ||||||                 :::   ::::::   :::
Db      3109 AGCTTTTGACTACAAAACTGAATAT---------------GACAACAAGCTGGGTAATTT 3153

Qy      1250 rLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrPr 1270
                |||   |||   ::::::                                 |||
Db      3154 AATGGACTACTATGGCATAAAAACAGAG------------------GCTGAAATACTTAG 3195

Qy      1270 oGluAlaMetAspLysSerGlyAlaAsnTyrAsp-----------------------Se 1282
                :::    |||:::         :::|||                            ||
Db      3196 TGGTGGCATTATGAAGGCATCAAAAACTTTTGACCGCAGAAAAGATGCTGAGGCCATTAG 3255

Qy      1282 rLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp-- 1301
             |     :::|||   |||                      :::|||         |||
Db      3256 TGTTGCTGTGAGGGCCTTGAGG------------------AAGGAGGCAAGAGCCTGGTT 3297

Qy      1302 ------------------------AlaLeuLeuArgAlaSerThrAlaPheLysLeuTy 1313
                                     |||   :::||||||      :::    :::
Db      3298 CAAGAGGCGTAATGATATAGATGACATGTTACCAAAGGCTTCGGCTTGGTACCACGTTAC 3357

Qy      1313 rTyrHisLysSer-----------------------------------------ProLy 1319
               ||||||   :::
Db      3358 ATATCATCCTACATATTGGGGTTGCTACAATCAGGGGTTGAAAAGAGCTCATTTCATTAG 3417

Qy      1319 sPheValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerAr 1339
               |||   |||   :::         :::   :::         |||   :::      :::||
Db      3418 CTTTCCCTGGTGTGTTTATGACCAGCTAATCCAGATTAAGAAGGACAAAGCACGTAACAG 3477

Qy      1339 gProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAs 1359
             ||||                     :::|||               :::      :::
Db      3478 GCCAGTTCTCAACTTGTCATCTCTCAGGGCTCAACTGAGTCACAGATTAGTG-------- 3529

Qy      1359 pLysLysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAs 1379
                                           |||:::   |||   |||
Db      3530 ------------------------------TTGAAATGAGATTCCAGTCGAGCGTTAAG 3558

Qy      1379 pProGluValTyrGluValLeuGlyAspAsp 1389
                :::|||   |||:::|||   |||
Db      3559 CTGATATATATATAATGTAATAGGGTGTGAT 3589
```

```
RESULT 11
AF468822
LOCUS       AF468822                6688 bp    DNA     linear   PLN 14-FEB-2002
DEFINITION  Diaporthe perjuncta putative RNA-dependent RNA polymerase RDP-1
            (rdp-1) gene, complete cds; and unknown genes.
ACCESSION   AF468822
VERSION     AF468822.1  GI:18656896
KEYWORDS    .
SOURCE      Diaporthe perjuncta
  ORGANISM  Diaporthe perjuncta
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Diaporthales; Valsaceae; Diaporthe.
REFERENCE   1  (bases 1 to 6688)
  AUTHORS   Preisig,O. and Wingfield,M.J.
  TITLE     Putative fungal RNA-dependent RNA polymerase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6688)
  AUTHORS   Preisig,O. and Wingfield,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2002) Forestry and Agricultural Biotechnology
            Institute, University of Pretoria, Lunnon Street 74, Pretoria 0002,
            South Africa
FEATURES             Location/Qualifiers
     source          1..6688
                     /organism="Diaporthe perjuncta"
                     /mol_type="genomic DNA"
                     /isolate="CMW3407"
                     /db_xref="taxon:186170"
                     /country="South Africa"
     mRNA            join(<1..275,348..>381)
                     /product="unknown"
     CDS             join(<1..275,348..381)
                     /note="similar to D-lactate dehydrogenase (D-lactate
                     ferricytochrome C oxidoreductase)"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAL78033.1"
                     /db_xref="GI:18656897"
                     /translation="MDDLGLFASILGHIGDGNFHESILYNPVNGEREKVEECVKKMVR
                     RAIDMDGTCTGEHSIGWGKKDSLLLEVGPETLAVMRAIKGSLDPRWIMNPGKIFDRQS
                     "
     gene            <1361..>5870
                     /gene="rdp-1"
     mRNA            join(<1361..4598,4654..>5870)
                     /gene="rdp-1"
                     /product="putative RNA-dependent RNA polymerase RDP-1"
     CDS             join(1361..4598,4654..5870)
                     /gene="rdp-1"
                     /codon_start=1
                     /product="putative RNA-dependent RNA polymerase RDP-1"
                     /protein_id="AAL78034.1"
                     /db_xref="GI:18656898"
                     /translation="MSSHHSTPTTPDRGPSAQPDASAQIHHTQTSSIPDATAKDNNPR
                     ISTLRDPAIADDTSDASNRNTQPSVVRRLPSPLSMSHTSNGNVLGQKVSLSGPIQGRS
                     IQTNASARTFGPRGHPNPRQNRFMPNRRQQRSPQSLQQRSTPAPLTGAVLSAAQSLTI
                     DLFGLPTDIKIREIWESMTSNHLNVNRVEIYENFDGKRTGKARVIIEPPPIRWPTWSG
                     PDNLITIDRADSSSLQVSLETRPFYPSTRTPCNRPIPAQFCQKKLLPTKLVFGMMASE
                     SAMLNMPQFSYEDVVLLTKFSKRQFELQFTVNLGRYGCKAFKSQIDFSHPTKIIWGRR
                     ENGAWALAMLLPSPPKLYGRLPVSHDDEQTLWQEWDLWARQTNISSEMHKLRDLPAKL
                     EKQTWKSIEFGRWTTYYLDMESSLSEWESVRTYLADYNVKILQVDHFDLVNPGLSNAW
                     KLLDNSAEMTLELLGSTQSYHLDFPVRYQLEVCISHGLLSEYGVNAQFLEKLISFETD
                     RARMMLEGVAEANKQFFEPMEMFDDPRILHYWPNARIPPYATLVRRAVITPTTIYFKT
                     PCVELTNRILRKYSDLNDRFLRVQFTDEVTFGKIFSSQDSNKDDNLYTRVHRVMQNGV
                     IIGDRHYKFLAFSNSQFRENGAFFFCETDHVTCASIRNWMGDFRHIRSVGKFAARMGQ
                     CFTTTRQVNGISIPKIRQIGDIQRRSGDNIWNFTDGVGKISVFFARMIASERDLPETP
                     SCFQMRMGGCKGVVVVWPDIPANEVHIRPSQEKFKAVYNGLEIIKTSTFAHATLNKQV
                     IPVLIALGVDDAVFVRMLDDELKEYDEALADSMKAGELLRSQVDENQTTLTMAEMVDT
                     FMDSEEPFLWTLLRLWKCWVLKRLKQKFAISVKNSAMIFGVVDETGVLRGHSQDTEGK
                     GFNNIESLPQIFLQVPIEGSDGKSTTNYEVITGICVVGRNPALHPGDVRVVEAVDVPE
                     LRHLKNVVVFPKTGDRDVPSMCSGGDMDGDDYFVYWDERLIPSEWDHPPLDHDADSST
                     PGLDSPADVTNGDVTRFFAQYMKNDSLGRIATAHFAQADQLAGGVKHPKCIELAKLHS
                     KAVDYIKSGKPAVMKRHLQPRNWPHWMERDKKSSYHSRGALGQIYDRIKIEEFHAAYE
                     MSFDTRILSRYQLEEDTLAKASEIKAAYDIAMRRLTGQHEAPVTEFEIWSTFILTKPR
                     VGSDYKLQENVGREMAALKARFRAVCMEAVTGDEQTPTCASASSKVNLEKLDRFVAAM
                     YTVTHNEVRAAVRERSKPRPDDEGDLEEIQMPLISFPWLYHRELARVALGRGGDVRPL
                     RKPVCADPAAANIESEPSGDGALSEEIDQEGETQLSLQKVRSKKDGYVIDRGESVETG
                     KVGEDYVRTSSGQVVRRGQLLTLFTDTGEEVQGAQSTIAPATPPAENNSGNPLVEDDG
                     TGSSYPASPRSEHEAEGEDEEIEFEEVDGANDGEEEDALEILARKIGM"
     mRNA            <6068..>6526
                     /product="unknown"
     CDS             6068..6526
```

```
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAL78035.1"
                     /db_xref="GI:18656899"
                     /translation="MVLQNHQQANRPTFLSLTYSSRASDYLISLVMAGDCFPLTWGPH
                     MWMQQPADQLGTSPGTPSRDDPFWHGQRDSQHRAFFQLAPGTSLDVLDWIGLESAEIP
                     ARLYSSLSSIHIRLSPSRPSTMDSRAVSSSLRTADDSQKVAHCQLFMLPF"
ORIGIN

Alignment Scores:
Pred. No.:              0.000468        Length:         6688
Score:                  296.50          Matches:        329
Percent Similarity:     33.38%          Conservative:   168
Best Local Similarity:  22.10%          Mismatches:     537
Query Match:            4.01%           Indels:         462
DB:                     8               Gaps:           69

US-09-913-878A-2 (1-1402) x AF468822 (1-6688)

Qy          6 ProArgLysArgAsnSerProValGluGluIleIleAsnArgLeuAsnAsnAspTyrAsn 25
                ||||||    :::       |||   :::       |||:::
Db       1452 CCTCGATTCCAGACGCAACCGCAAAAGACAACAATCCACGAATTTCTACCCTTCGTGACC 1511

Qy         26 LeuGlyLeuGlnCysValAlaAspThrThrLeuThrProHisArgArgLysGluLeuAla 45
              |||  ||||||                       ||||||    |||       |||:::
Db       1512 CTGCTATTGCAGATGACACATCTGATGCGTCTAACCGCAACACGCAACCATCCGTTGTCC 1571

Qy         46 GluSerAspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyr 65
                       |||||||||                  :::                  |||
Db       1572 ---------GAAGACTTCCGTCACCGCTGAGCATGTCACACACCTCCAATGGCAACGTAC 1622

Qy         66 TrpArgLysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAla 85
              |||    :::      ||||||:::                  |||:::     |||
Db       1623 TGGGCCAGAAGGTC-AGCCTATCCGGACCTATTCAAGGTCGCTCGATTCAGACCAATGCG 1681

Qy         86 AlaSerSerAsnTrpValProLysAlaHisAlaAspProAspThrLeuProTrpSerLys 105
              ::::::      :::    |||:::   |||     :::|||
Db       1682 TCAGCTCGGACATTTGGTCCTCGAGGTCATCCCAATCCACGGCAG--------------- 1726

Qy        106 GluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeuGlu 125

Db       1726 ------------------------------------------------------------ 1726

Qy        126 ValLeuAsnArgPheMetPro---------------------------------ProPro 134
                    ||||||||||||||
Db       1727 ------AATAGATTCATGCCTAATCGGAGACAGCAGCGTTCACCGCAGAGCTTGCAACAG 1780

Qy        135 AsnAsnThrProGlyArgThrPheGlyArgThrLeuSerGlyProSerGlyLeuSer--- 153
                 :::||||||         |||      ||||||         |||:::
Db       1781 CGCAGCACACCTGCTCCATTGACGGGTGCGGTCCTAAGCGCCGCTCAGAGCTTGACCATC 1840

Qy        154 ------------ArgProThrSer------------------------------------ 157
                           ||||||
Db       1841 GACTTGTTTGGACTTCCAACAGATATCAAGATTCGCGAAATATGGGAGTCGATGACTTCC 1900

Qy        158 ------ThrAsnThrLysArgLysAspGluProAlaAsnValThrPheAlaAspProPro 175
                       |||      |||   :::         |||            |||
Db       1901 AACCATCTCAACGTCAACAGAGTCGAGATCTATGAGAATTTC---------GATGGAAAG 1951

Qy        176 LysArgSerLeuThrArgSerAlaThrGlyProProIleHisGlyAlaAlaIleProLeu 195
              :::               |||            ||||||               |||:::
Db       1952 AGGACAGGGAAAGCGAGAGTAATCATCGAGCCACCC------------------CCGATC 1993

Qy        196 LysPhePro--------------------------------AspProValAsnThrGly 204
              ::::::|||                                 |||      ::::::
Db       1994 AGATGGCCAACATGGAGCGGACCTGACAACCTCATCACCATTGATAGAGCAGACTCCTCG 2053

Qy        205 SerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArgAlaLysGly 224
              |||   :::   |||||||||:::                     |||:::
Db       2054 AGTCTGCAGGTGTCTCTGGAGACGAGACCATTCTACCCATCCACCCGAACGCCGTGCAAC 2113

Qy        225 LysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValProIleAlaSerAlaLeu 244
              :::                                       |||:::|||
Db       2114 AGG-------------------------------CCAATCCCGGCCCAGTTCTGCCA- 2139

Qy        245 AspLysValProThrArgArgHisAlaAsnThrArgAspProThrAlaThrGlyHis--- 263
              ::::::    |||       :::            |||           ||||||
Db       2140 GAAGAAACTCCTGCCCACAAAGTTGGTGTTTGGAATGATGGCCAGCGAGTCGGCCATGTT 2199

Qy        264 -----------------------------ArgArgAlaAspGlnValAspSerPheAsp 273
                                          |||      |||:::|||          |||
Db       2200 GAACATGCCACAGTTCAGCTATGAGGACGTCGTCTTGTTGACAAAGTTCT-CCAAACGAC 2258

Qy        274 ThrSerGlnGlyThrSerTyrGlySerSerValPheSerAlaCysArgHisAsnGlnSer 293
                       :::|||         |||      :::|||   :::|||
Db       2259 AATTTGAACTTCAGTTCACGGTCAATCTCGGCCGATACGGCTGCAAAGCAT--------- 2309

Qy        294 ThrThrGlnSerSerPheGluAlaProProSerGlnProArgGluLysArgProValAsp 313
                 :::   |||      |||         :::|||   |||                |||
Db       2310 TTAAATCGCAAATCGATTTCAGCC------ACCCAACGAAGATCATCTGGGGTCGTCGCG 2363

Qy        314 AlaThr-ValPheGluAlaGlyHis-LeuIleGluSerProSerLys------GlyArgT 331
                 |||          ||||||  |||:::      ||||||    |||       ||||||
Db       2364 AGAACGGAGCATGGGCTTTGGCCATGTTGCTACCCAGTCCTCCCAAACTCTATGGTCGTT 2423

Qy        331 hrThrLysSerHisIleAsp-AsnGlnProLeuSerSerSerSerGlnGlyGluThrSer 350
                    ||||||  |||  |||:::|||           :::
Db       2424 TGCCTGTCTCACATGATGACGAACAAACCCTGTGGCAGGAAT------------------ 2465

Qy        351 PheSerThrTyrTyrGluSerPheProSerSerGlyGlyGlu------------GlyAla 366
                   ||||||             |||:::|||                     |||
Db       2466 ---GGGACTTATGGGCCCGACAGACCAACATCTCGAGTGAGATGCACAAGTTGAGGGACC 2522

Qy        367 IleProGluProSerArgSerAsnGlyLeuAlaArgSerGluGluSerAlaArgSerGln 386
                 |||   ||||||         :::                  ||||||      |||
Db       2523 TTCCCG---CCAAGCTTGAGAAGCAAACCTGGAAATCCATCGAATTCGGCCGCTGGACAA 2579

Qy        387 ValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLysPhePro 406
              :::   :::                                    :::   |||
Db       2580 CTTATTATC----------------------------TCGACATGG------------ 2597

Qy        407 LysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPheMet--- 425
                          ||||||   |||   ||||||         |||      :::::
Db       2598 ---------------AATCCTCTCTGAGTGAG-TGGGAGAGTGTCAGAACGTACTTGGCA 2641

Qy        426 ------HisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSer 443
                    :::   |||:::      |||:::                        :::|||
Db       2642 GATTACAACGTCAAGATC---TTGCAAGTCGACCATTTCGATCTC---GTCAATCCTGGT 2695

Qy        444 TrpSerThrAlaArgAspValThrAspIleTrpLysThrLeu------------------ 457
                 |||   |||                     ||||||   |||
Db       2696 CTCTCCAATGCC------------------TGGAAACTGCTCGACAATTCGGCCGAGATG 2737

Qy        458 ------------------------------TyrArgLeuAspAlaPheArgGlyLysPro 467
                                            |||      ||||||
Db       2738 ACACTCGAGCTCCTCGGTTCAACACAGAGTTACCACCTTGAC------------------ 2779

Qy        468 PhePro--------------------------------------------GluLysPro 472
              ||||||                                             |||
Db       2780 TTTCCAGTGCGGTATCAGTTGGAAGTCTGCATCTCTCATGGCCTCCTCAGCGAATACGGT 2839

Qy        473 ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSerLysGlySerAlaVal 492
                 |||      |||:::      :::      :::|||||||:::      :::   :::
Db       2840 GTCAATGCCCAGTTTCTCGAAAAGCTCATC---AGCTTCGAGACCGACAGAGCCAGAATG 2896

Qy        493 ValLeuSerAlaValLeuAspTyrAsnPro--------------------------Asp 503
              :::|||      |||   :::   |||                               |||
Db       2897 ATGCTCGAAGGTGTCGCTGAAGCTAACAAGCAGTTCTTTGAGCCCATGGAGATGTTCGAT 2956

Qy        504 AsnSer------------------ProThrAla------ProLeuTyrLeuValLysLeu 515
              :::                      |||   |||        |||   |||      :::
Db       2957 GACCCGAGAATACTGCATTACTGGCCAAATGCGAGGATCCCCCCTTACGCAACTCTGGTC 3016

Qy        516 Lys---------------------ProLeuMetPheGluGlnGlyCys----------- 524
              :::                           :::   |||:::      |||
Db       3017 CGCAGGGCTGTTATCACTCCCACCACAATCTATTTCAAGACTCCTTGCGTCGAGCTGACC 3076
```

```
Qy    525 ---ArgLeuThrArgArgPheGlyPro------AspArgPhe-PheGlu----------- 537
Db   3077 AATCGAATCCTCCGCAAGTACTCTGACCTGAACGACCGGTTCCTTCGAGTCCAATTCACT 3136

Qy    538 -IleLeuIleProSerProThrSerThrSerProSerValProProValValSer----- 555
Db   3137 GACGAGGTTACCTTTGGCAAAATCTTTTCATCCCAGGATTCCAACAAAGATGACAACCTG 3196

Qy    556 -------LysGlnProGlyAlaValGluGluValIle----------------------- 565
Db   3197 TACACCCGCGTACACAGGGTCATGCAAAACGGTGTTATCATCGGGGACCGTCATTATAAA 3256

Qy    566 -GlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPhePheAl 585
Db   3257 TTCTTGGCGTTCA---GCAACTCACAGTTTA---GGGAGAACGGC-GCATTCTTCTTTTG 3309

Qy    585 aLysAsp---------------AlaGlyTyrArgLysProLeuArgGluPheGlnLeuAr 600
Db   3310 TGAGACCGATCATGTCACTTGTGCCTCCATCCGCAACTGGATGGGGGACTTCCGG----- 3364

Qy    600 gAlaGluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIl 620
Db   3365 ------------------CATATCCGCTCCGTCGGCAAGTTTGCTGCGAGAATGGGTCA 3405

Qy    620 eThrPheArgProAspValPheLysThrArgSerValValProAlaGluGluProValGl 640
Db   3406 GTGTTTCACC------------ACAACACGACAAGTCAATGGCATCAGCATACCA----- 3448

Qy    640 uGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAspAsnAs 660
Db   3449 ----------------AAAATTCGCCAGATTGGTGATATACAGCGGCGGAGTGGTGACAA 3492

Qy    660 nThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTy 680
Db   3493 CATCTGGAAC-------------------------------------------------- 3502

Qy    680 rAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAspLeuLeuSerPr 700
Db   3502 ------------------------------------------------------------ 3502

Qy    700 oSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSerValAlaLy 720
Db   3503 ------------------TTCACTGATGGAGTGGCAAGATATCCGTCTTCTTTGCAAG 3543

Qy    720 sArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGl 740
Db   3544 AATGATTGCAAGCGAGCGTGATTTGCCAGAGACGCCTTCATGCTTTCAGATGAGAATGGG 3603

Qy    740 ySerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGl 760
Db   3604 CGGGTGCAAGGGCGTGGTCGTTGTATGGCCTGACATCCCAGCGAACGAG------GTCCA 3657

Qy    760 uThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLe 780
Db   3658 TATACGACCTTCTCAGGAGAAGTTTAAGGCCGTGTAC--------------AATGGTCT 3702

Qy    780 uGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuPr 800
Db   3703 CGAGATCATCAAG---ACCTCTACGTTCGCCCATGCTACTCTCAATAAGCAGGTTATACC 3759

Qy    800 oValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLe 820
Db   3760 TGTCCTGATCGCCCTGGGTGTGGAT---------------GACGCCGTTTTCGTCAGAAT 3804

Qy    820 uIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGl 840
Db   3805 GCTCGATGATGAGCTCAAAGAGTACGACGAA---------GCCTTA------------- 3841

Qy    840 uPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyAr 860
Db   3842 ---------------------GCTGATTCAATGAAAGCCGGTGAACTTCTTCGGTCTCA 3879

Qy    860 gValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeu---Me 879
Db   3880 AGTC--------------------GACGAGAACCAAACAACATTGACCATGGCGGAGAT 3918

Qy    879 tAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGl 899
Db   3919 GGTCGACACGTTCATGGACTCCGAAGAGCCGTTCCTC-------------TGGACCTTGCT 3966

Qy    899 nLysArg---LysCys---------AspThrLeuLysSerLysLeuAsnIleArgValGl 915
Db   3967 TCGTTTGTGGAAGTGCTGGGTGTTGAAGAGGTTAAAGCAAAAGTTCGCCATCAGCGTCAA 4026

Qy    915 yArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsn----- 933
Db   4027 GAACAGCGCTATGATCTTCGGGGTTGTCGATGAGACCGGTGTGCTGAGGGGCCACTCACA 4086

Qy    934 -------------------------------------------------------GluVa 935
Db   4087 AGACACCGAAGGCAAAGGGTTCAATAACATCGAGTCGCTCCCACAGATCTTTCTCCAAGT 4146

Qy    935 lHisVal---GlyPheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeuSe 954
Db   4147 CCCTATCGAGGGCAGTGATGGGAAGAGCACCACGAATTACGAGGTCATCACTGGAATT-- 4204

Qy    954 rAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgVa 974
Db   4205 ----------TGCGTAGTAGGCCGTAACCCAGCGCTTCATCCTGGAGATGTGCGGGTTGT 4254

Qy    974 lArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLy 994
Db   4255 TGAGGCAGTTGACGTGCCAGAGTTACGACATCTCAAGAACGTGGTCGTCTTTCCAAAGAC 4314

Qy    994 sGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTr 1014
Db   4315 TGGAGACAGGGACGTTCCGAGCATGTGCTCCGGTGGTGACATGGACGGAGATGACTACTT 4374

Qy   1014 pValCysTrpAspProGluIleVal---------------------------AspGlyPh 1025
Db   4375 TGTTTACTGGGACGAGCGACTCATCCCAAGCGAGTGGGACCACCCACCTCTGGACCACGA 4434

Qy   1025 eValAsnAlaGluMetPro------------------------LeuGluProAspLeuSe 1037
Db   4435 TGCGGATTCATCGACTCCTGGGCTGGACAGCCCGGCAGATGTGACAAACGGGGACGTCAC 4494

Qy   1037 rArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGl 1057
Db   4495 TAGGTTC--------------------TTTGCCCAGTATATGAAGAACGATTCCCTTGG 4533

Qy   1057 ySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLe 1077
Db   4534 CCGTATCGCC---------------------------------ACAGCGCACTTTGCTCA 4560

Qy   1077 uGlnProAsnPheLeuGly---------------MetCysThrAsnTyrLysGluArgLe 1092
Db   4561 GGCAGATCAGTTGGCAGGGGGCGTCAAGCACCCAAAATGTAAGTAAATCGCTACGACCTT 4620

Qy   1092 uCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGl 1112
Db   4621 GTGCTGGC--CGGGGTGAACTGACGGCTGAAGTAGGCATCGAGCTCGCCAAGCTCCACTC 4678

Qy   1112 yAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLe 1132
Db   4679 AAAAGCTGTCGACTATATCAAGTCCGGC-----------------AAACCCGCGGTCAT 4720

Qy   1132 uArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyr------LysSe 1150
Db   4721 GAAGAGACATCTT------CAGCCTCGAAACTGGCCGCATTGGATGGAGAGGGACAAGAA 4774

Qy   1150 rAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIl 1170
Db   4775 GTCCTCCTACCACTCCAGAGGCGCGCTGGGTCAAATTTACGACCGGATCAAG-------- 4826

Qy   1170 eAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLy 1190
```

```
                               :::|||  ||||||   |||   :::   :::
Db     4827 ------------------------ATCGAGGAATTCCATGCCGCTTACGAAATGTCTTT 4861

Qy     1190 sAspThr--------------------------GluAspGlyAlaHisPheTrpAspPr 1201
             ||||||                           ||||||
Db     4862 CGACACGCGCATCCTATCGCGCTACCAGCTGGAGGAAGATACTTTGGCGAAAGCCAGCGA 4921

Qy     1201 oAspLeuAlaSerTyrTyrThrPhePheLysGluIleSer-AspLysSerArg------S 1219
                     |||:::|||       :::  ::::::  |||   :::|||       |
Db     4922 GATAAAAGCCGCCTACGACATCGCTATGCGCCGTCTGACGGGACAACACGAGGCACCTGT 4981

Qy     1219 erSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrG 1239
                 ||  :::|||  |||      |||   :::                      |
Db     4982 CACGGAGTTTGAGATTTGGTCGACTTTCATCC---------TCACCAAGCCTCGCGTTGG 5032

Qy     1239 lyArgLeuValLysAsnLysGluMetArgAspSerLysAspProTyrProValArgValA 1259
             ||   ::::::   :::|||  |||          :::   |||:::       |||   :
Db     5033 GAGCGATTATAAGCTCCAAGAGAATGTTGGCCGGGAGATGGCCGCAT---TGAAGGCCCG 5089

Qy     1259 snGlnValTyrGluLysTrpCysAlaIleThrProGluAlaMetAspLysSerGlyAlaA 1279
             ::            |||   |||:::|||   :::               :::   |||:
Db     5090 ATTCCGCGCGGTGTGCATGGAG-GCCGTGACTGGGGATGAGCAGACCCCAACTTGCGCTT 5148

Qy     1279 snTyrAspSerLysVal 1284
             ::      |||||||||
Db     5149 CTGCCAGTTCCAAGGTC 5165
```

```
RESULT 12
DDI314909
LOCUS       DDI314909            5731 bp    DNA     linear   INV 14-JUN-2001
DEFINITION  Dictyostelium discoideum partial RrpA gene for putative RNA
            dependent RNA polymerase, exons 1-2.
ACCESSION   AJ314909
VERSION     AJ314909.1  GI:14475568
KEYWORDS    RNA dependent RNA polymerase; RrpA gene.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1
  AUTHORS   Martens,H., Novotny,J., Oberstrass,J., Steck,T.L., Postlethwait,P.
            and Nellen,W.
  TITLE     RNAi in Dictyostelium: developmental regulation and the role of
            RdRPs and dsRNase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5731)
  AUTHORS   Martens,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2001) Martens H., Genetics Department, Kassel
            University, Heinrich-Plett Strasse 40, 34132, GERMANY
FEATURES             Location/Qualifiers
     source          1. .5731
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
     gene            1. .5731
                     /gene="RrpA"
     mRNA            join(<1. .1572,1655. .>5731)
                     /gene="RrpA"
     CDS             join(1. .1572,1655. .>5731)
                     /gene="RrpA"
                     /function="required for RNA interference"
                     /codon_start=1
                     /product="putative RNA dependent RNA polymerase"
                     /protein_id="CAC41974.1"
                     /db_xref="GI:14475569"
                     /db_xref="GOA:Q95ZG7"
                     /db_xref="SPTREMBL:Q95ZG7"
                     /translation="METKDVHYKPTKLPQEKPMIPKFKEGVEIIPRDYQIESYYQSTQ
                     DNTLLVLPTGMGKTLVSIMTLLEMFSINDQDNSCGDSKRIALFLVDRVPLVTQQAGAI
                     EAITNLKVCKLYGEINDSRTRAFVRSKEYDVLVSTVGSLVNLLEVRHLNILDFYFITF
                     DEVHHATGEHDFNKVVDYIRKTDLNYRPRILGLTASLVSIGNSTIDIVQRSIKDMEER
                     MLSRVFKPTSLLTNSTQSELQPELVSFKTSGQESLIETSICEFLVSNSKELQIFLNYS
                     DVDLDGVRGNPSFLKALERLDKYSKQLESIYIEYTKVLIELYALLSVLSTEGPKQVLT
                     RLDLLMQSTNQENELYNKLLELRYYISGQYEDRESNFEKGSTRYRKLISSLEYAISDA
                     SDGELQKDLRILVFVETRFGASNLTSMLKKEPFQEYLHTKRLVGHNGVDGMDSEKQQS
                     IIRKFRDGKCRLIVTTNVLEEGIDVQDCNIVICYDGILSLKSLIQRRGRARSKNESKF
                     IIIYNDDKDLKILDILSSENLLNNSINQILSNRERQAITKNQFLNWKKLNINNEYFTN
                     NQLDGGPLDKDIKCVILSIYHIEIDIETFRELLLFQLPDIINQVDQVSIRMEGDDEYL
                     SSYIIVILNTCIDKEQSIYRLKSLLNINGLKYWMIDRSSFHFEDDNINDLSQSKNKNY
                     SIRSFEYGNMSSPTKYHGIGQLDDSFELFYGKREMFFTKKTRFQRYRLNLKLTDIEST
                     FLLVNDKYQNIVHILLIVKRPLIFEEEKKTHIVFWNKSKPKKDDFGQAFVYRLTLSTI
                     DNEIKEVIKPLIANGFRMYESSLELNCFIQGGINKNNNCMIIDLNLQYIKRNLEIHYL
                     LQCLQSQSSYGRVLNQDFIVEINNLLAFGKQEEAEFIITSLISKRTKFINYKEFIGDC
                     LKNFLPDPYYAKKKDLSTKSGEFTMIRDVFITPSRTIFIQPTLQRSCRAVRKFGSSNF
                     IVVKIVNESLEPLQGVNNKHPLHENRIKPILEGGILVGGEIYSYAGNSNSQLREYSSW
                     FVSNQIGTHTVKIWSGIEHVDNVRKFFRCIGLMFSTTIPTVTLPQNRIYRIQDITRNT
                     HEFTEGCGEIGPELAKHLNENYNFRPSTCAYQVRIGGNKGVLVVNNQAPDPSGIYIRP
                     SMVKFNPIDCGDEHRTLEICSVSTTSRCKLNRQVISLLSTLGTQDNVFFALQDHYLNQ
                     VAQIVNDTNASKQAIVEFFPDITEGELYQDPYIRRILISLYKLKMERIQQKCHIEIKD
                     SRMLLGVCDPTNSLPPNTVFVQLEEEDEDDDDDGRKYEKVIEGLVMVIKNPCTHPGDV
                     RYLKAVDNIRLRHLRNVLVFSTKGDVPNFKEISGSDLDGDRYFFCYDKSLIGNRSESE
                     TAYLVVETVSNNDKKANVFNDPFALSSMYSTNAERQELGKLYHSHLAISDQYGANHKY
                     SIQISKECFKEIDYPKTGIHGTIPKEVNIRLQTVGYPHYMQRENSTRVYYQSKKIMGK
                     MYDQIDQLVYIGDFLPNISLDKSNLVDGYEIYLNSAKILYSQYKLQVHSLLRHYSAES
                     EESIMIGFLDQGFISDKVSKDIKGEMKNDYIKIQQTFENEFLKEFGEQHKENCLLIHR
                     VNIEKKVSAWYHVAYSDLKDDRALGFYWIARSFKQMTLDQESQKKKNILVNSILDHMN
                     KKKDTLFPLYESRLLIVKELTSFLNTLNDTVFKESTLEIIGSTATMLFDENSNLNLYL
                     KLENQNDHMNLAEKNEILNQLKDEIKTSSFISAEACSIIKIQGDIPLFKFKIDKIQCL
                     IGLETSKWYETLVQQQYITKNPSILSILFSIIDWSINSLFVDSETGEIINNSKVYSSL
                     NLQKVIVENSMIPSIFNNKNHNNTTIIIIIIIIIIIIIII"
     exon            <1. .1572
                     /gene="RrpA"
                     /number=1
     intron          1573. .1654
                     /gene="RrpA"
                     /number=1
     exon            1655. .>5731
                     /gene="RrpA"
                     /number=2
ORIGIN

Alignment Scores:
Pred. No.:              0.000522        Length:         5731
Score:                  293.50          Matches:        173
Percent Similarity:     37.68%          Conservative:   145
Best Local Similarity:  20.50%          Mismatches:     305
Query Match:            3.97%           Indels:         221
DB:                     3               Gaps:           41

US-09-913-878A-2 (1-1402) x DDI314909 (1-5731)

Qy      509 ProLeuTyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrArg 528
            |||         :::  ::::::|||   :::   ::::::  ||||||        |||
Db     2921 CCAAGTAGAACTATTTCATTCAACCAACACTA---CAAAGATCATGTAGAGCTGTTCGT 2977

Qy      529 ArgPheGlyProAspArgPhePheGluIleLeuIleProSerProThrSerThrSerPro 548
            :::||||||        |||      :::   |||                     :::
Db     2978 AAATTTGGTTCATCCAATTTCATTGTGGTTAAAATT---------------GTAAATGAA 3022

Qy      549 SerValProProValValSerLysGlnProGlyAlaValGluGluValIleGlnTrpLeu 568
            |||:::   |||:::
Db     3023 TCATTGGAACCACTT--------------------------------------------- 3037

Qy      569 ThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPhePheAlaLysAspAla 588
                                                                  :::
Db     3038 ---------------------------------------------------CAAGGTGTA 3046

Qy      589 GlyTyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysProIleIle 608
               :::   ||||||   |||   ::::::                  |||||||||:::
Db     3047 AATAATAAACATCCACTTCATGAGAATAGAATT---------------AAACCAATACTT 3091

Qy      609 LysGluArgValHisPhePheAlaGluThrGlyIleThrPheArgProAspValPheLys 628
                                     |||   ||||||         :::::::::
Db     3092 -----------------------GAGGGTGGTATATTGGTTGGTGGTGAAATCTATTCG 3127
```

```
Qy    629 ThrArgSerValValProAlaGluGluProValGluGlnArgThrGluPheLysValSer 648
Db   3128 TATGCTGGT-----------------------------AATAGTAATAGTCAGTTAAGA 3157
Qy    649 GlnMetLeuAspTrpLeuLeu------GlnLeuAspAsnAsnThr---------TrpGln 663
Db   3158 GAATATTCATCTTGGTTCGTTTCTAATCAAATTGGTACACATACGGTTAAAATTTGGAGT 3217
Qy    664 ------ProHisLeu------------LysLeuPheSerArgIleGlnLeuGlyLeuSer 677
Db   3218 GGTATAGAGCATGTTGATAATGTTAGAAAGTTTTTCAGATGTATTGGTTTAATGTTTTCA 3277
Qy    678 LysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAspLeu 697
Db   3278 ACAACGATTCCAACAGTTACATTACCACAGAATCGTATC---TATAGAATCCAAGATATC 3334
Qy    698 LeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSer 717
Db   3335 ------ACCAGAAATACCCATGAA---TTCACCGAGGGATGTGGTGAAATTGGTCCAGAA 3385
Qy    718 ValAlaLysArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGly 737
Db   3386 CTTGCTAAACATTTAAATGAAAATTATAATTTCAGACCATCAACATGTGCATATCAAGTT 3445
Qy    738 ArgPheGlySerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAsp 757
Db   3446 AGAATAGGTGGAAATAAAGGTGTGTTGGTGGTAAAC---AATCAAGCACCCGATCCATCT 3502
Qy    758 TrpIleGluThrTyrProSerGlnArgLysTrpGlu---CysAspPheValAspLysHis 776
Db   3503 GGTATTTATATTAGACCAAGTATGGTGAAATTCAATCCAATCGATTGTGGTGATGAACAT 3562
Qy    777 GlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeu 796
Db   3563 ---AGAACATTAGAGATTTGCAGTGTT---TCGACAACTTCAAGATGTAAATTAAATAGA 3616
Qy    797 GlnLeuLeuProValLeuGluAsp---ArgAlaArgAspLysValLysMetArgGlnAla 815
Db   3617 CAAGTCATCAGTTTACTCTCAACATTGGGTACACAAGATAATGTT--------------- 3661
Qy    816 IleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeu 835
Db   3662 --------------------------------TTCTTTGCATTACAAGATCATTATCTA 3688
Qy    836 AsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArg 855
Db   3689 AAT---------CAAGTGGCTCAAATCGTAAATGATACCAATGCTTCCAAACAAGCAATC 3739
Qy    856 ValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeu 875
Db   3740 GTTGAATTC-----------------------TTTCCAGATATTACAGAGGGTGAACTT 3775
Qy    876 AsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAla 895
Db   3776 -----------------------------TATCAAGATCCTTACATTCGTAGAATTTTA 3805
Qy    896 TrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGly 915
Db   3806 ATTAGTCTTTATAAATTAAAAATGGAGAGAATTCAACAAAAATGTCATATAGAGATTAAA 3865
Qy    916 ArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluVal 935
Db   3866 GATAGTAGAATGTTGTTGGGTGTTTGTGATCCAACAAATTCATTACCACCAAATACTGTA 3925
Qy    936 HisValGlyPheSerSerLysPheArgAspGluGluGlu---------SerPheThrLeu 952
Db   3926 TTTGTTCAATTGGAAGAGGAAGATGAAGATGATGATGATGATGGTCGTAAATATGAAAAA 3985
Qy    953 LeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGln 972
Db   3986 GTAATCGAAGGATTGGTAATGGTAATTAAAAATCCTTGTACTCATCCTGGTGATGTAAGA 4045
Qy    973 ArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSer 992
Db   4046 TATCTAAAGGCTGTTGATAATATAAGATTAAGACATTTAAGAAATGTATTGGTTTTCTCA 4105
Qy    993 ThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMet 1012
Db   4106 ACAAAAGGTGACGTACCAAACTTTAAAGAAATTTCAGGTTCAGATTTAGATGGTGATCGT 4165
Qy   1013 AlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeu 1032
Db   4166 TATTTCTTTTGTTACGATAAATCATTGATTGGTAATCGTTCAGAATCAGAG--------- 4216
Qy   1033 GluProAspLeuSerArgTyrLeu-----------------------LysLysAspLys 1044
Db   4217 ------------ACAGCCTATTTAGTTGTCGAGACCGTTTCAAACAATGATAAAAAGGCC 4264
Qy   1045 ThrThrPheLysGln---------LeuMetAlaSerHisGlyThrGlySerAlaAlaLys 1061
Db   4265 AATGTTTTCAATGATCCATTTGCTTTATCCTCAATGTACTCTACAAATGCTGAACGTCAA 4324
Qy   1062 GluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPhe 1081
Db   4325 GAA------------------CTTGGTAAATTGTATCAT---------------TCACAT 4351
Qy   1082 LeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsn 1101
Db   4352 TTGGCAATCTCTGATCAATAC--------------------------GGTGCAAATCAC 4384
Qy   1102 LysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGly 1121
Db   4385 AAATATTCAATTCAAATCTCTAAAGAATGTTTTAAAGAAATTGATTATCCAAAAACTGGT 4444
Qy   1122 IleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeu 1141
Db   4445 ATTCATGGCACTATTCCAAAAGAAGTAAACATTCGGTTACAAACTGTAGGT--------- 4495
Qy   1142 SerLeuProAspProMet--------------------------TyrLysSerAspSer 1152
Db   4496 ---TATCCACATTATATGCAAAGAGAGAATTCAACCAGAGTCTATTATCAATCGAAAAAA 4552
Qy   1153 TrpLeuGlyArg---------------GlyGluProThrHisIleIleAspTyrLeu--- 1166
Db   4553 ATTATGGGTAAAATGTATGATCAAATTGATCAGTTGGTTTATATTGGTGATTTCTTACCA 4612
Qy   1167 LysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMet 1186
Db   4613 AATATTAGTCTTGATAAAAGTAATTTGGTTGATGGCTATGAAATCTATTTAAATAGTGCT 4672
Qy   1187 LysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyr 1206
Db   4673 AAAATC--------------------------------------TTATATAGTCAG 4690
Qy   1207 Tyr--------------ThrPhePheLysGluIleSerAspLysSerArgSerSerAla 1221
Db   4691 TATAAACTTCAAGTTCATTCATTGTTAAGACATTATTCAGCAGAATCTGAAGAATCTATA 4750
Qy   1222 LeuLeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeu 1241
Db   4751 ATG-----------------------ATTGGTTTCCTTGATCAAGGTTTC--------- 4777
Qy   1242 ValLysAsnLysGluMetArgAsp-----------SerLysAspProTyrProValArg 1257
Db   4778 ATTAGTGATAAAGTTTCAAAAGATATCAAAGGTGAAATGAAAAATGATTAC---ATTAAA 4834
Qy   1258 ValAsnGlnValTyrGluLysTrpCysAlaIleThrProGluAlaMetAspLysSerGly 1277
Db   4835 ATTCAACAAACTTTTGAAAAT------------------GAGTTTTTAAAGGAATTTGGT 4876
Qy   1278 AlaAsnTyrAspSerLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGlu 1297
Db   4877 GAACAACACAAAGAGAATTGTTTATTGATTCATAGAGTAAACATT------GAAAAGAAG 4930
Qy   1298 MetAsnThrTrp 1301
```

```
Db       4931 GTTTCAGCTTGG 4942

RESULT 13
AY148431
LOCUS       AY148431              3807 bp    mRNA    linear   PLN 12-NOV-2002
DEFINITION  Arabidopsis thaliana RNA-dependent RNA polymerase 1 mRNA, complete
            cds.
ACCESSION   AY148431
VERSION     AY148431.1  GI:24935274
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 3807)
  AUTHORS   Yu,D., Fan,B., MacFarlane,S.A. and Chen,Z.
  TITLE     Analysis of the involvement of an inducible Arabidopsis
            RNA-dependent RNA polymerase in antiviral defense
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3807)
  AUTHORS   Yu,D., Fan,B., MacFarlane,S.A. and Chen,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-SEP-2002) Botany and Plant Pathology, Purdue
            University, West Lafayette, IN 47907, USA
FEATURES             Location/Qualifiers
     source          1..3807
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
     CDS             128..3451
                     /note="RdRP1"
                     /codon_start=1
                     /product="RNA-dependent RNA polymerase 1"
                     /protein_id="AAN64409.1"
                     /db_xref="GI:24935275"
                     /translation="MGKTIQVFGFPNGVSAEEVKKFLERLTGSGTVYAIKVRQPKKGG
                     PRVYAIVQFTSERHTRLIITAAAERLYYGRSYLKAFEVEQDIVPKPRASLHTISGLKM
                     FFGCQVSTKKFLTLWSAQDVCVSFGIGMRKLHFSFSWYQKDYRLELSYENIWQIDLHS
                     PQGRSSKFLVIQVIGAPKIFEKEDQPINLLFGIMDFYSDGSDEQWIRTTDFTSSSCIG
                     QSTAFCLELPVHLNVPDFRENFANYAEHRASSFLIESGSSYSSNANTLVPVVDPPPGF
                     SLPFEILFKLNTLVQNACLSGPALDLDFYRLLNQKKYDRALIDHCLEKLFHLGECCYE
                     PAHWLRDEYKKWISKGKLPLSPTISLDDGLVYMYRVQVTPARVYFSGPEVNVSNRVLR
                     HYSKYINNFLRVSFVDEDLEKVRSMDLSPRSSTQRRTKLYDRIYSVLRDGIVIGDKKF
                     EFLAFSSSQLRENSAWMFAPIDRITAAHIRAWMGDFDHIRNVAKYAARLGQSFSSSRE
                     TLNVRSDEIEVIPDVEIISLGTRYVFSDGIGKISAEFARKVARKCGLTEFSPSAFQIR
                     YGGYKGVVAVDPNSSKKLSLRKSMSKFESENTKLDVLAWSKYQPCYMNRQLITLLSTL
                     GVTDSVFEKKQREVVDRLDAILTHPLEAHEALGLMAPGENTNILKALILCGYKPDAEP
                     FLSMMLQNFRASKLLELRTKTRIFISGGRSMMGCLDETRTLEYGQVVVQYSDPMRPGR
                     RFIITGPVVVAKNPCLHSGDVRVLQAVNVPALNHMVDCVVFPQKGLRPHPNECSGSDL
                     DGDIYFVCWDQELVPPRTSEPMDYTPEPTQILDHDVTIEEVEEYFANYIVNDSLGIIA
                     NAHTAFADKEPLKAFSDPCIELAKKFSTAVDFPKTGVAAVIPQHLYVKEYPDFMEKPD
                     KPTYESKNVIGKLFREVKERAPPLISIKSFTLDVASKSYDKDMEVDGFEEYVDEAFYQ
                     KANYDFKLGNLMDYYGIKTEAEILSGGIMRMSKSFTKRRDAESIGRAVRALRKETLSL
                     FNASEEEENESAKASAWYHVTYHSSYWGLYNEGLNRDHFLSFAWCVYDKLVRIKKTNL
                     GRRQRQETLERLDHVLRFG"
ORIGIN

Alignment Scores:
Pred. No.:              0.00044        Length:         3807
Score:                  290.00         Matches:        174
Percent Similarity:     38.02%         Conservative:   129
Best Local Similarity:  21.83%         Mismatches:     302
Query Match:            3.92%          Indels:         195
DB:                     8              Gaps:           33

US-09-913-878A-2 (1-1402) x AY148431 (1-3807)

Qy        614 PhePheAlaGluThrGlyIleThrPheArgProAspValPheLysThr-ArgSerValVa 633
Db       1396 TTCTACTCAAAGAAGAACCAAGTTATACGACAGGATATATTCTGTTCTTCGCGATGGGAT 1455

Qy        633 lProAlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAsp-- 652
Db       1456 TGTCATAGGCGATAAGAAGTTTGAGTTTCTCGCCTTTTCTTCCAGCCAGTTGCGGGAGAA 1515

Qy        653 -------TrpLeuLeu---GlnLeuAspAsnAsnThr-----------------TrpGl 663
Db       1516 CTCTGCTTGGATGTTTGCACCAATAGATCGGATTACAGCAGCACATATCAGAGCTTGGAT 1575

Qy        663 n---------ProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTy 680
Db       1576 GGGTGATTTTGACCATATAAGAAATGTAGCTAAATATGCTGCAAGGCTTGGTCAATCTTT 1635

Qy        680 rAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAspLeuLeuSer-- 699
Db       1636 TAGC---TCTTCAAGAGAGACACTTAATGTTAGGAGTGATGAGATTGAAGTGATTCCTGA 1692

Qy        700 ------------ProSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSe 715
Db       1693 TGTTGAGATTATATCTTTAGGGACACGCTATGTGTTTTCCGATGGAATAGGAAAGATATC 1752

Qy        715 rArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal---ProSerAl 734
Db       1753 AGCTGAATTTGCTAGAAAGGTAGCGAGGAAATGTGGTCTTACAGAGTTTTCTCCATCTGC 1812

Qy        734 aValGlnGlyArgPheGlySerAlaLysGlyMet-TrpValIleAspValAspAspThrG 754
Db       1813 TTTTCAAATCCGTTATGGCGGTTATAAAGGAGTGGTGGCTGTTG--------------- 1856

Qy        754 lyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAspPheValA 774
Db       1857 --------------------------------------------------------ATCC 1860

Qy        774 spLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyL 794
Db       1861 AAACTCATCAAAGAAA-CTGTCTCTGAGGAAGAGTATGAGCAAATTCGAATCGGAGAACA 1919

Qy        794 euAsnLeuGlnLeuLeu-----------------ProValLeuGluAspArgAlaArgA 808
Db       1920 CCAAGCTTGATGTTCTGGCGTGGAGCAAGTACCAACCTTGTTATATGAACAGACAACTGA 1979

Qy        808 spLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnP 828
Db       1980 TTACACTTTTGTCTACTCTTGGAGTTACGGACAGTGTGTTTGAGAAGAAACAAAGGGAAG 2039

Qy        828 heSerGluGlnLysHisAla---LeuAsnArgProValGluPheArgGlnTrpValTyrG 847
Db       2040 TTGTGGATCGTCTCGACGCAATCTTGACTCATCCTTTGGAA------------------ 2080

Qy        847 luSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyL 867
Db       2081 -----------------------------GCTCACGAGGCTCTTGGTTTAATGGCT---- 2107

Qy        867 euProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysL 887
Db       2108 --CCAGGGGAAAACACAAATATTCTCAAGGCATTGATCTTGTGTGGCTATAAACCTGACG 2165

Qy        887 ysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuL 907
Db       2166 CTGAACCTTTCCTTTCAATGATGCTTCAGAATTTCAGGGCATCTAAGTTGTTGGAACTAC 2225

Qy        907 ysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheT 927
Db       2226 GGACAAAAACAAGGATTTTCATTTCTGGTGGAAGATCTATGATGGGATGCCTAGACGAGA 2285

Qy        927 rpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluG 947
Db       2286 CCAGAACGCTGGAGTATGGTCAGGTAGTAGTGCAGTATTCAGATCCCATGAGGCCGGGAA 2345

Qy        947 luGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisP 967
Db       2346 GGCGA---------TTCATCATCACCGGACCTGTTGTTGTTGCCAAAAACCCATGCCTGC 2396

Qy        967 heProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysA 987
```

```
Db     2397 ATTCTGGTGACGTGCGTGTTCTTCAAGCTGTCAATGTCCCAGCTTTAAATCATATGGTGG 2456

Qy      987 spValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyA 1007
Db     2457 ACTGCGTTGTGTTCCCGCAAAAGGGCTTGAGGCCACACCCAAATGAATGTTCTGGGAGTG 2516

Qy     1007 spTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValA 1027
Db     2517 ATTTAGATGGAGATATTTACTTTGTATGTTGGGATCAAGAATTGGTTCCGCCAAGAACGT 2576

Qy     1027 snAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrP 1047
Db     2577 CTGAACCAATGGACTACACTCCTGAACCAACTCAAATCTTGGATCATGAT---------- 2626

Qy     1047 heLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrA 1067
Db     2627 --------------------------------------------------GTCACAATTG 2636

Qy     1067 spMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrA 1087
Db     2637 AGGAAGTTGAAGAGTACTTTGCGAACTACATTGTGAATGATAGTTTAGGGATCATCGCGA 2696

Qy     1087 snTyr------------------LysGluArgLeuCysTyrIleAsnAsnSerValSerA 1101
Db     2697 ATGCTCATACCGCCTTTGCGGATAAGGAACCACTC---------------AAGGCGTTTA 2741

Qy     1101 snLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnG 1121
Db     2742 GTGACCCATGCATTGAGCTTGCAAAGAAGTTTTCAACTGCCGTAGACTTTCCAAAAACTG 2801

Qy     1121 lyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeu-----------L 1137
Db     2802 GTGTTGCA------------------GCTGTGATACCGCAACATCTTTACGTGAAAGAGT 2843

Qy     1137 euGlyGlyAlaLeuSerLeuProAsp---ProMetTyrLysSerAspSerTrpLeuGly- 1155
Db     2844 ATCCTGATTTCATGGAGAAACCGGATAAGCCAACATATGAGTCGAAGAACGTGATTGGTA 2903

Qy     1156 ----------------------ArgGlyGluProThrHisIleIleAspTyrLeuLysP 1168
Db     2904 AGCTTTTTAGAGAGGTAAAAGAGCGAGCTCCACCGTTGATCTCGATCAAATCATTCACGC 2963

Qy     1168 heSerIleAlaArgProAlaIleAspLysGluLeu------------------------ 1179
Db     2964 TTGATGTGGCCTCAAAGTCTTATGATAAAGACATGGAAGTTGACGGATTTGAGGAGTATG 3023

Qy     1180 -----GluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisP 1198
Db     3024 TTGATGAAGCTTTCTAC------CAGAAGGCGAATTATGATTTCAAGTTAGGT------- 3070

Qy     1198 heTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGlu--------------- 1212
Db     3071 ----------AATTTGATGGATTACTATGGGATTAAGACTGAGGCTGAGATTCTTAGTG 3119

Qy     1213 -----IleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgI 1231
Db     3120 GTGGTATAATGAGGATGTCAAAGTCA-----------TTCACC------AAGAGACGAG 3161

Qy     1231 leGlyGlu---Val-GluLysGluTyrGlyArg-------------------------- 1240
Db     3162 ATGCAGAATCGATTGGAAGAGCGGTTCGGGCGTTGAGGAAAGAAACTTTGTCGTTGTTCA 3221

Qy     1241 ---LeuValLysAsnLysGluMetArgAspSerLysAspPro------TyrProValArg 1257
Db     3222 ATGCTTCTGAAGAAGAAAATGAG-TCAGCAAAGGCTTCGGCTTGGTATCATGTAACG 3280

Qy     1258 ValAsnGlnValTyrGluLysTrpCysAlaIleThrProGluAlaMetAspLysSerGly 1277
Db     3281 TACCACTCAAGTTAC------TGG-----------------GGACTTTATAATGAGGGT 3316

Qy     1278 AlaAsnTyrAsp------------------------------SerLysValIleArgLeu 1287
Db     3317 TTGAATCGTGACCATTTCTTAAGTTTTGCGTGGTGCGTTTATGATAAGCTTGTGAGGATC 3376

Qy     1288 LeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrpAlaLeuLeuArgAlaSer 1307
Db     3377 AAGAAGACCAATCTTGGGAGGCGTCAGCGGCAGGAGACGCT-CGAGCGGTTAGACCATGT 3435

Qy     1308 ThrAlaPheLysLeuTyrTyrHisLysSerProLysPheValTrp 1322
Db     3436 TCTGCGTTTCGGTTGATTTTT---------------TTTTTTTGG 3465

RESULT 14
AF239718
LOCUS       AF239718                4013 bp    DNA     linear   PLN 02-JUN-2000
DEFINITION  Arabidopsis thaliana SGS2 gene, complete cds.
ACCESSION   AF239718
VERSION     AF239718.1  GI:8164027
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 4013)
  AUTHORS   Mourrain,P., Beclin,C., Elmayan,T., Feuerbach,F., Godon,C.,
            Morel,J.B., Jouette,D., Lacombe,A.M., Nikic,S., Picault,N.,
            Remoue,K., Sanial,M., Vo,T.A. and Vaucheret,H.
  TITLE     Arabidopsis SGS2 and SGS3 genes are required for
            posttranscriptional gene silencing and natural virus resistance
  JOURNAL   Cell 101 (5), 533-542 (2000)
  MEDLINE   20306668
   PUBMED   10850495
REFERENCE   2  (bases 1 to 4013)
  AUTHORS   Beclin,C., Mourrain,P., Vaucheret,H. and Elmayan,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2000) Biologie Cellulaire, INRA, Route de
            Saint-Cyr, Versailles 78026, France
FEATURES             Location/Qualifiers
     source          1..4013
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="III"
     mRNA            join(<1..2715,3138..>4013)
                     /product="SGS2"
     CDS             join(1..2715,3138..4013)
                     /note="RNA-dependent RNA polymerase-like protein"
                     /codon_start=1
                     /product="SGS2"
                     /protein_id="AAF73959.1"
                     /db_xref="GI:8164028"
                     /translation="MGSEGNMKKSVVTQVSIGGFGESTTAKQLTDYLEDEVGIVWRCR
                     LKTSWTPPGSYPNFEIADTSNIPSIDEYKKVEPHAFVHFAVFESAGRAMDAAGQCNLI
                     LDGQPLKVSLGPKNPYSLNQRRRTTVPYKLAGITLEIGTLVSRDDFFVSWRAEGVDFL
                     VDPFDNTCKFCFRKSTAFSFKDAVMHAVINCDYKLELLVRDIQTVRQYKTLHGFVLIL
                     QLASSPRVWYRTADDDIYDTVPGDLLDDDDPWIRTTDFTQVGAIGRCHSYRVLISPRY
                     ENKLRTALDYFRMRRVQEERVRWPPRIRNEPCFGEPVSDHFFCIHHKEGISFEIMFLV
                     NSVLHRGVFNQFQLTERFFDLLRNQPKDVNIASLKHLCTYKRPVFDAYKRLKLVQEWI
                     QKNPKLLGSHEQSEDISEIRRLVITPTRAYCLPPEVELSNRVLRRYKAVAERFLRVTF
                     MDESMQTINSNVLSYFVAPIVKDLTSSSFSQKTYVFKRVKSILTDGFKLCGRKYSFLA
                     FSANQLRDRSAWFFAEDGKTRVSDIKTWMGKFKDKNVAKCAARMGLCFSSTYATVDVM
                     PHEVDTEVPDIERNGYVFSDGIGTITPDLADEVMEKLKLDVHYSPCAYQIRYAGFKGV
                     VARWPSKSDGIRLALRDSMKKFFSKHTILEICSWTRFQPGFLNRQIITLLSVLGVPDE
                     IFWDMQESMLYKLNRILDDTDVAFEVLTASCAEQGNTAAIMLSAGFKPKTEPHLRGML
                     SSVRIAQLWGLREKSRIFVTSGRWLMGCLDEAGILEHGQCFIQVSKPSIENCFSKHGS
                     RFKETKTDLEVVKGYVAIAKNPCLHPGDVRILEAVDVPQLHHMYDCLIFPQKGDRPHT
                     NEASGSDLDGDLYFVAWDQKLIPPNRKSYPAMHYDAAEEKSLGRAVNHQDIIDFFARN
                     LANEQLGTICNAHVVHADRSEYGAMDEECLLLAELAATAVDFPKTGKIVSMPFHLKPK
                     LYPDFMGKEDYQTYKSNKILGRLYRRVKEVYDEDAEASSEESTDPSAIPYDAVLEIPG
                     FEDLIPEAWGHKCLYDGQLIGLLGQYKVQKEEEIVTGHIWSMPKYTSKKQGELKERLK
                     HSYNSLKKEFRKVFEETIPDHENLSEEEKNILYEKKASAWYHVTYHPEWVKKSLELQD
                     PDESSHAAMLSFAWIAADYLARIKIRSREMGSIDSAKPVDSLAKFLAQRL"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                 0.000496      Length:         4013
Score:                     289.50        Matches:        190
Percent Similarity:        35.43%        Conservative:   120
Best Local Similarity:     21.71%        Mismatches:     297
Query Match:               3.91%         Indels:         269
DB:                        8             Gaps:           41

US-09-913-878A-2 (1-1402) x AF239718 (1-4013)

Qy    535 PhePheGluIleLeuIleProSerProThrSerThrSerProSerValProProValVal 554
Db   1396 TTTGTTGCTCCTATTGTGAAGGATTTGACATCAAGTTCTTTCTCCCAGAAGACCTACGTT 1455

Qy    555 SerLysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSer 574
Db   1456 TTTAAAAGAGTGAAGAGCATA------------------TTAACCGATGGG---TTTAAA 1494

Qy    575 LeuValGlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeu 594
Db   1495 CTATGTGGTAGAAAATAC---AGTTTTCTAGCATTCTCAGCCAAT--------------- 1536

Qy    595 ArgGluPheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhe 614
Db   1537 ---------CAACTGAGA-----------------------GACCGCTCTGCATGGTTC 1563

Qy    615 PheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValPro 634
Db   1564 TTTGCTGAAGACGGG-----------------------AAAACACGT------------ 1587

Qy    635 AlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeu 654
Db   1588 ------------------------------------GTGTCAGATATAAAAACATGGATG 1611

Qy    655 LeuGlnLeu---AspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGln 673
Db   1612 GGGAAGTTCAAAGACAAGAATGTG------------GCAAAATGTGCTGCTAGGATGGGC 1659

Qy    674 LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHis 693
Db   1660 CTGTGCTTCTCCTCCACATATGCCACTGTAGATGTCATGCCTCACGAGGTT--------- 1710

Qy    694 LysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArg 713
Db   1711 GACACTGAGGTTCCAGATATTGAGAGAAATGGGTATGTTTTCTCTGACGGAATTGGTACA 1770

Qy    714 MetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal------ 731
Db   1771 ATCACACCTGACCTCGCTGACGAAGTAATGGAGAAACTTAAGTTG---GATGTGCACTAC 1827

Qy    732 ---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspVal 750
Db   1828 AGCCCTTGTGCTTATCAGATACGTTACGCAGGTTTCAAAGGGGTT--------------- 1872

Qy    751 AspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln--------------- 765
Db   1873 -----------------------GTTGCTCGTTGGCCATCAAAAAGTGATGGAATCAGG 1908

Qy    766 ---------ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluVal 782
Db   1909 CTAGCCCTTCGAGACAGTATGAAGAAGTTCTTTTCCAAACAT---ACGATCTTGGAGATC 1965

Qy    783 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 802
Db   1966 TGTTCC---TGGACGAGGTTTCAACCTGGGTTCTTAAATCGGCAGATAATTACCCTTCTA 2022

Qy    803 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 822
Db   2023 TCC-------------------------GTACTAGGTGTTCCGGATGAAATATTCTGG 2055

Qy    823 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArg 842
Db   2056 GATATGCAG---------GAATCCATGCTCTATAAACTGAACCGCATCCTTGATGATACA 2106

Qy    843 GlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValPro 862
Db   2107 GATGTGGCATTTGAAGTTCTCACGGCATCATGTGCT------------------------ 2142

Qy    863 PheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGly 882
Db   2143 ---------------------GAACAGGGAAACACTGCAGCTATCATGCTTAGTGCAGGT 2181

Qy    883 PheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLys 902
Db   2182 TTCAAACCAAAAACCGAGCCGCATCTACGCGGGATGTTGTCTTCAGTCAGAATTGCACAA 2241

Qy    903 CysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMet 922
Db   2242 CTCTGGGGTCTCAGAGAAAAATCTCGTATTTTTGTTACTTCAGGAAGGTGGCTAATGGGT 2301

Qy    923 IleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer------ 940
Db   2302 TGCCTAGACGAAGCAGGGATACTTGAACATGGCCAATGCTTTATTCAAGTCTCTAAACCG 2361

Qy    941 ------------------------------SerLysPheArgAspGluGluGluSerPhe 950
Db   2362 TCTATAGAAAATTGTTTCTCCAAACATGGGTCTCGTTTTAAGGAGACAAAGACAGATCTG 2421

Qy    951 ThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAsp 970
Db   2422 GAAGTAGTTAAAGGCTAT---GTAGCCATTGCTAAGAATCCTTGTCTTCACCCAGGGGAT 2478

Qy    971 IleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIle 990
Db   2479 GTAAGGATTTTAGAAGCTGTTGATGTACCCCAGCTGCATCACATGTATGACTGCCTTATT 2538

Qy    991 PheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGly 1010
Db   2539 TTCCCTCAGAAAGGTGATAGGCCGCATACAAACGAAGCTTCTGGCAGTGACCTTGACGGG 2598

Qy   1011 AspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet 1030
Db   2599 GACCTGTACTTTGTGGCTTGGGATCAGAAACTCATC------------------------ 2634

Qy   1031 ProLeuGluProAspLeuSerArgTyr------LeuLysLysAspLysThrThrPheLys 1048
Db.  2635 ------CCTCCCAACAGGAAAAGCTATCCGGCCATGCATTATGATGCAGCTGAAGAGAAG 2688

Qy   1049 GlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyr------ 1066
Db   2689 AGTTTA-----------------GGCCGTGCTGTCAACCACCAGGTCAA-TATCTACTT 2729

Qy   1067 --------------------------------------AspMetIleGlnLysSerPhe 1073
Db   2730 CCCCACTTTTTTTCTGTTAGCTGAGATGATGACGCTTAAGATAAAATGGACAAGTGATTC 2789

Qy   1074 HisPhe--------------------AlaLeuGlnProAsnPheLeuGlyMetCysThr 1086
Db   2790 CAGTATAACTCTCTGATATTTCTGGAAGTGGTGGAAACTTTGTTTTTGGATGCTTGTAGT 2849

Qy   1087 AsnTyr------------------------------------LysGluArgLeuCysTyr 1094
Db   2850 GATTACTTTTGCTTGCTTCTGAGTTCATAGATTGCGTGGCATAGGTTGTTAACAGCATAC 2909

Qy   1095 IleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeu 1114
Db   2910 TGTGTTCATCCTGTAGGAGCTAAG---ATGATGACATTTAAGAAAACTTTGTCATATATT 2966

Qy   1115 ValAspGlnSer-------------------------------------------Lys 1119
Db   2967 GTTGATTCTGATTTGTATTTTTCGCATGCCTTACTGTGACTGTGACTGTTTAGTTACTAT 3026

Qy   1120 GlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu------------ 1135
Db   3027 CAAGGGCAGTATCATGATGCTAGTAGCTTCTGCCAATGTAGTTATGAATTCTTTGTTATT 3086
```

```
Qy     1136 --------------------LeuLeuGlyGly---AlaLeuSerLeuProAspProMet 1147
                                :::|||||         |||   |||    |||
Db     3087 TTTTCTTGCTGTGCTTGCCAAGTCTTGGGTAATTGATTGCTTTTGTTACAGGAC------ 3140

Qy     1148 TyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLys 1167
                                                     |||||||||:::
Db     3141 -----------------------------------------ATAATCGATTTC------ 3152

Qy     1168 PheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLys 1187
                  ||||||          :::::::::|||                |||
Db     3153 ------TTTGCAAGAAACTTGGCGAATGAGCAGTTGGGCACAATTTGCAAT--------- 3197

Qy     1188 AlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyr 1207
                                    ||||||           |||   :::   |||
Db     3198 ----------------------GCACACGTCGTTCATGCTGATAGAAGTGAGTATGGA 3233

Qy     1208 ThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeu 1227
                     :::|||                         ||||||         |||
Db     3234 GCCATGGACGAAGAA-----------------------TGTTTGCTACTGGCAGAACTA 3269

Qy     1228 LysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuVal--------------- 1242
                              :::      |||      |||::::::|||
Db     3270 GCTGCCACTGCAGTTGATTTCCCAAAGACA---GGGAAAATTGTGTCAATGCCCTTCCAC 3326

Qy     1243 ---LysAsnLysGluMetArgAsp------SerLysAspProTyrProVal--------- 1256
               |||   |||          |||         |||:::   |||
Db     3327 CTAAAACCAAAACTCTACCCAGATTTCATGGGAAAAGAAGACTACCAAACTTACAAGTCG 3386

Qy     1257 ---------------:-----------ArgValAsnGlnValTyrGluLysTrpCys--- 1266
                                       ||||||   :::||||||::::::
Db     3387 AACAAAATCTTGGGTCGGCTTTACAGACGGGTAAAAGAGGTTTATGATGAAGATGCAGAA 3446

Qy     1267 AlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysValIleArg 1286
            |||   :::   |||:::   |||   |||         ||||||:::
Db     3447 GCTTCCTCAGAAGAAAGCACAGACCCAAGTGCCATCCCTTATGACGCT------------ 3494

Qy     1287 LeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp 1301
            :::||||||:::                |||            |||
Db     3495 GTTCTTGAAATACCGGGATTTGAAGATTTGATCCCTGAGGCATGG 3539
```

RESULT 15
AX204839

```
LOCUS       AX204839                6863 bp    DNA     linear   PAT 30-AUG-2001
DEFINITION  Sequence 1 from Patent WO0155407.
ACCESSION   AX204839
VERSION     AX204839.1  GI:15394183
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Beclin,C., Elmayan,T., Mourrain,P. and Vaucheret,H.
  TITLE     Novel sgs2 plant gene and use thereof
  JOURNAL   Patent: WO 0155407-A 1 02-AUG-2001;
            Rhobio (FR) ; INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (FR)
FEATURES             Location/Qualifiers
     source          1..6863
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN
```

Alignment Scores:			
Pred. No.:	0.000998	Length:	6863
Score:	289.50	Matches:	190
Percent Similarity:	35.43%	Conservative:	120
Best Local Similarity:	21.71%	Mismatches:	297
Query Match:	3.91%	Indels:	269
DB:	6	Gaps:	41

US-09-913-878A-2 (1-1402) x AX204839 (1-6863)

```
Qy      535 PhePheGluIleLeuIleProSerProThrSerThrSerProSerValProProValVal 554
            |||             ::::::            ||||||:::|||   |||         |||
Db     2245 TTTGTTGCTCCTATTGTGAAGGATTTGACATCAAGTTCTTTCTCCCAGAAGACCTACGTT 2304

Qy      555 SerLysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSer 574
               |||:::      ::::::                  ||||||   |||
Db     2305 TTTAAAAGAGTGAAGAGCATA------------------TTAACCGATGGG---TTTAAA 2343

Qy      575 LeuValGlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeu 594
            |||   ||||||::::::   :::|||      |||
Db     2344 CTATGTGGTAGAAAATAC---AGTTTTCTAGCATTCTCAGCCAAT--------------- 2385

Qy      595 ArgGluPheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhe 614
                     |||||||||                                          |||
Db     2386 ---------CAACTGAGA-----------------------GACCGCTCTGCATGGTTC 2412

Qy      615 PheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValPro 634
            |||||||||   |||                        |||||||||
Db     2413 TTTGCTGAAGACGGG-----------------------AAAACACGT------------ 2436

Qy      635 AlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeu 654
                                                    ||||||   :::      |||:::
Db     2437 ------------------------------------GTGTCAGATATAAAAACATGGATG 2460

Qy      655 LeuGlnLeu---AspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGln 673
               :::         |||   |||                  |||      :::|||:::
Db     2461 GGGAAGTTCAAAGACAAGAATGTG------------GCAAAATGTGCTGCTAGGATGGGC 2508

Qy      674 LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHis 693
            |||      |||   |||||||||   :::   :::   ||||||::::::
Db     2509 CTGTGCTTCTCCTCCACATATGCCACTGTAGATGTCATGCCTCACGAGGTT--------- 2559

Qy      694 LysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArg 713
              |||::::::                        |||      |||      :::||||||:::|||
Db     2560 GACACTGAGGTTCCAGATATTGAGAGAAATGGGTATGTTTTCTCTGACGGAATTGGTACA 2619

Qy      714 MetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal------ 731
            ::::::      :::|||      :::   :::   |||   |||   ||||||
Db     2620 ATCACACCTGACCTCGCTGACGAAGTAATGGAGAAACTTAAGTTG---GATGTGCACTAC 2676

Qy      732 ---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspVal 750
               |||   |||   |||   |||:::            ||||||:::
Db     2677 AGCCCTTGTGCTTATCAGATACGTTACGCAGGTTTCAAAGGGGTT--------------- 2721

Qy      751 AspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln--------------- 765
                                   :::      :::||||||:::
Db     2722 ----------------------GTTGCTCGTTGGCCATCAAAAAGTGATGGAATCAGG 2757

Qy      766 ---------ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluVal 782
                     |||               |||      ||||||         ||||||:::
Db     2758 CTAGCCCTTCGAGACAGTATGAAGAAGTTCTTTTCCAAACAT---ACGATCTTGGAGATC 2814

Qy      783 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 802
               |||      :::      :::         ||||||   |||::::::   :::|||
Db     2815 TGTTCC---TGGACGAGGTTTCAACCTGGGTTCTTAAATCGGCAGATAATTACCCTTCTA 2871

Qy      803 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 822
                                                   :::   |||   :::
Db     2872 TCC--------------------------GTACTAGGTGTTCCGGATGAAATATTCTGG 2904

Qy      823 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArg 842
            |||:::|||                  :::      ||||||||   ::::::
Db     2905 GATATGCAG---------GAATCCATGCTCTATAAACTGAACCGCATCCTTGATGATACA 2955

Qy      843 GlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValPro 862
                     :::|||      ::::::
Db     2956 GATGTGGCATTTGAAGTTCTCACGGCATCATGTGCT----------------------- 2991

Qy      863 PheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGly 882
                                 |||      |||               ::::::::::|||
Db     2992 ---------------------GAACAGGGAAACACTGCAGCTATCATGCTTAGTGCAGGT 3030
```

```
Qy    883 PheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLys 902
Db   3031 TTCAAACCAAAAACCGAGCCGCATCTACGCGGGATGTTGTCTTCAGTCAGAATTGCACAA 3090

Qy    903 CysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMet 922
Db   3091 CTCTGGGGTCTCAGAGAAAAATCTCGTATTTTTGTTACTTCAGGAAGGTGGCTAATGGGT 3150

Qy    923 IleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer------ 940
Db   3151 TGCCTAGACGAAGCAGGGATACTTGAACATGGCCAATGCTTTATTCAAGTCTCTAAACCG 3210

Qy    941 ------------------------------SerLysPheArgAspGluGluGluSerPhe 950
Db   3211 TCTATAGAAAATTGTTTCTCCAAACATGGGTCTCGTTTTAAGGAGACAAAGACAGATCTG 3270

Qy    951 ThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAsp 970
Db   3271 GAAGTAGTTAAAGGCTAT---GTAGCCATTGCTAAGAATCCTTGTCTTCACCCAGGGGAT 3327

Qy    971 IleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIle 990
Db   3328 GTAAGGATTTTAGAAGCTGTTGATGTACCCCAGCTGCATCACATGTATGACTGCCTTATT 3387

Qy    991 PheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGly 1010
Db   3388 TTCCCTCAGAAAGGTGATAGGCCGCATACAAACGAAGCTTCTGGCAGTGACCTTGACGGG 3447

Qy   1011 AspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet 1030
Db   3448 GACCTGTACTTTGTGGCTTGGGATCAGAAACTCATC------------------------ 3483

Qy   1031 ProLeuGluProAspLeuSerArgTyr------LeuLysLysAspLysThrThrPheLys 1048
Db   3484 ------CCTCCCAACAGGAAAAGCTATCCGGCCATGCATTATGATGCAGCTGAAGAGAAG 3537

Qy   1049 GlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyr------ 1066
Db   3538 AGTTTA------------------GGCCGTGCTGTCAACCACCAGGTCAA-TATCTACTT 3578

Qy   1067 ------------------------------------------AspMetIleGlnLysSerPhe 1073
Db   3579 CCCCACTTTTTTCTGTTAGCTGAGATGATGACGCTTAAGATAAAATGGACAAGTGATTC 3638

Qy   1074 HisPhe---------------------AlaLeuGlnProAsnPheLeuGlyMetCysThr 1086
Db   3639 CAGTATAACTCTCTGATATTTCTGGAAGTGGTGGAAACTTTGTTTTTGGATGCTTGTAGT 3698

Qy   1087 AsnTyr------------------------------------LysGluArgLeuCysTyr 1094
Db   3699 GATTACTTTTGCTTGCTTCTGAGTTCATAGATTGCGTGGCATAGGTTGTTAACAGCATAC 3758

Qy   1095 IleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeu 1114
Db   3759 TGTGTTCATCCTGTAGGAGCTAAG---ATGATGACATTTAAGAAAACTTTGTCATATATT 3815

Qy   1115 ValAspGlnSer---------------------------------------------Lys 1119
Db   3816 GTTGATTCTGATTTGTATTTTTCGCATGCCTTACTGTGACTGTGACTGTTTAGTTACTAT 3875

Qy   1120 GlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu------------ 1135
Db   3876 CAAGGGCAGTATCATGATGCTAGTAGCTTCTGCCAATGTAGTTATGAATTCTTTGTTATT 3935

Qy   1136 ---------------------LeuLeuGlyGly---AlaLeuSerLeuProAspProMet 1147
Db   3936 TTTCTTGCTGTGCTTGCCAAGTCTTGGGTAATTGATTGCTTTTGTTACAGGAC------ 3989

Qy   1148 TyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLys 1167
Db   3990 -----------------------------------------ATAATCGATTTC------ 4001

Qy   1168 PheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLys 1187
Db   4002 ------TTTGCAAGAAACTTGGCGAATGAGCAGTTGGGCACAATTTGCAAT--------- 4046

Qy   1188 AlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyr 1207
Db   4047 -----------------------GCACACGTCGTTCATGCTGATAGAAGTGAGTATGGA 4082

Qy   1208 ThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeu 1227
Db   4083 GCCATGGACGAAGAA------------------------TGTTTGCTACTGGCAGAACTA 4118

Qy   1228 LysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuVal--------------- 1242
Db   4119 GCTGCCACTGCAGTTGATTTCCCAAAGACA---GGGAAAATTGTGTCAATGCCCTTCCAC 4175

Qy   1243 ---LysAsnLysGluMetArgAsp------SerLysAspProTyrProVal--------- 1256
Db   4176 CTAAAACCAAAACTCTACCCAGATTTCATGGAAAAGAAGACTACCAAACTTACAAGTCG 4235

Qy   1257 --------------------------ArgValAsnGlnValTyrGluLysTrpCys--- 1266
Db   4236 AACAAAATCTTGGGTCGGCTTTACAGACGGGTAAAAGAGGTTTATGATGAAGATGCAGAA 4295

Qy   1267 AlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysValIleArg 1286
Db   4296 GCTTCCTCAGAAGAAAGCACAGACCCAAGTGCCATCCCTTATGACGCT------------ 4343

Qy   1287 LeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp 1301
Db   4344 GTTCTTGAAATACCGGGATTTGAAGATTTGATCCCTGAGGCATGG 4388

Search completed: March 31, 2004, 10:57:01
Job time : 9235.43 secs
```

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         March 31, 2004, 02:00:42 ; Search time 833.391 Seconds
                                           (without alignments)
                                           7146.671 Million cell updates/sec

Title:          US-09-913-878A-2
Perfect score:  7397
Sequence:       1 MNPITPRKRNSPVEEIINRL..........YEVLGDDDFDGIGFTGNGDY 1402

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:      6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09913878/runat_29032004_082007_8357/app_query.fasta_1.2254
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878_@CGN_1_1_797_@runat_29032004_082007_8357 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      N_Geneseq_29Jan04:*
                1:  geneseqn1980s:*
                2:  geneseqn1990s:*
                3:  geneseqn2000s:*
                4:  geneseqn2001as:*
                5:  geneseqn2001bs:*
                6:  geneseqn2002s:*
                7:  geneseqn2003as:*
                8:  geneseqn2003bs:*
                9:  geneseqn2003cs:*
                10:  geneseqn2004s:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	7397	100.0	8045	3	AAA65171	Aaa65171 Neurospor
	2	341	4.6	615	7	ABZ54843	Abz54843 Aspergill
	3	328	4.4	3591	4	AAH77704	Aah77704 Nucleotid
	4	319.5	4.3	3901	3	AAC63740	Aac63740 Maize RNA
	5	317	4.3	2816	3	AAC63741	Aac63741 Rice RNA-
	6	297.5	4.0	3731	4	AAS17845	Aas17845 Tomato RN
	7	297.5	4.0	3731	4	AAS17837	Aas17837 Tomato RN
	8	297.5	4.0	3731	4	AAD04370	Aad04370 Tomato RN
	9	289.5	3.9	6863	4	AAH77703	Aah77703 Nucleotid
	10	283	3.8	3807	3	AAC63742	Aac63742 Soybean R
	11	267.5	3.6	3737	3	AAC63739	Aac63739 Maize RNA
	12	236	3.2	1281	3	AAC63738	Aac63738 Maize RNA
	13	195.5	2.6	4551	7	ACA42620	Aca42620 Prokaryot
	14	192	2.6	49377	2	AAV05287	Aav05287 The sorap
	15	188.5	2.5	4148	3	AAZ90112	Aaz90112 Human ant
	16	188	2.5	4010	6	ABI99508	Abi99508 Mouse isc
	17	185.5	2.5	5922	4	ABL18812	Abl18812 Drosophil
	18	185	2.5	5710	4	ABL18813	Abl18813 Drosophil
	19	184	2.5	3824	4	AAI57835	Aai57835 Human pol
	20	183	2.5	3813	3	AAZ90111	Aaz90111 Human ant
	21	178.5	2.4	28958	2	AAT89956	Aat89956 Sorangium
	22	178.5	2.4	28958	3	AAA75299	Aaa75299 DNA seque
	23	176	2.4	4668	7	ABT19128	Abt19128 Aspergill
	24	176	2.4	4733	7	ABT18534	Abt18534 Aspergill
	25	176	2.4	5499	7	ABT20948	Abt20948 Aspergill
	26	176	2.4	5550	7	ABT20350	Abt20350 Aspergill
	27	176	2.4	6733	7	ABT17940	Abt17940 Aspergill
	28	176	2.4	7550	7	ABT19754	Abt19754 Aspergill
	29	175	2.4	13329	5	AAS71562	Aas71562 DNA encod
	30	174	2.4	2469	7	ACA38023	Aca38023 Prokaryot
	31	174	2.4	14770	6	ABL61893	Abl61893 Colon ade
	32	174	2.4	14770	6	ABL68257	Abl68257 Kidney ca
	33	174	2.4	14770	6	ABL68538	Abl68538 Kidney ca
	34	174	2.4	14770	9	ADB31314	Adb31314 Testoster
c	35	174	2.4	14797	5	AAS65159	Aas65159 DNA encod
	36	174	2.4	15187	5	AAS86131	Aas86131 DNA encod
c	37	174	2.4	36535	7	AAL56705	Aal56705 Chimpanze
	38	173.5	2.3	6814	4	ABL17599	Abl17599 Drosophil
	39	173.5	2.3	9059	4	ABL17598	Abl17598 Drosophil
	40	173.5	2.3	10074	6	ABK13416	Abk13416 Drosophil
	41	173.5	2.3	28598	2	AAT06769	Aat06769 Sorangium
	42	172	2.3	1353	8	ACF06132	Acf06132 Bacterial
c	43	171	2.3	52297	2	AAT51411	Aat51411 Mycobacte
c	44	171	2.3	52298	2	AAQ47357	Aaq47357 L5 mycoba
c	45	171	2.3	113193	7	AAD54645	Aad54645 Streptomy

ALIGNMENTS

```
RESULT 1
AAA65171
ID   AAA65171 standard; DNA; 8045 BP.
XX
AC   AAA65171;
XX
DT   16-NOV-2000  (first entry)
XX
DE   Neurospora crassa qde-1 gene.
XX
KW   Gene silencing; quelling deficient; qde-1; al-1; ds.
XX
OS   Neurospora crassa.
XX
FH   Key             Location/Qualifiers
FT   CDS             2447..6655
FT                   /*tag=  a
FT                   /product= "QDE-1"
XX
PN   WO200050581-A2.
XX
PD   31-AUG-2000.
XX
PF   16-FEB-2000; 2000WO-IT000048.
XX
PR   22-FEB-1999;   99IT-RM000117.
XX
PA   (UYRO-) UNIV ROMA LA SAPIENZA.
XX
PI   Macino G,  Cogoni C;
XX
DR   WPI; 2000-579171/54.
```

```
DR   P-PSDB; AAB13956.
XX
PT   Novel polynucleotide encoding a polypeptide which has a silencing
PT   activity and comprising a RNA-dependent RNA polymerase domain.
XX
PS   Claim 1; Page 31-43; 48pp; English.
XX
CC   The present sequence is the Neurospora crassa qde-1 gene. This gene has
CC   silencing activity. The qde-1 gene was isolated by mutational analysis of
CC   an al-1 transgenic strain. This strain had an albino phenotype resulting
CC   from post-transcriptional silencing of the endogenous al-1 gene.
CC   Reversion of this phenotype indicated a mutation in a silencing gene. The
CC   silencing gene, qde-1, could then be isolated. Modulation of qde-1
CC   expression may be used to inactivate genes and to silence suppression of
CC   genes
XX
SQ   Sequence 8045 BP; 2089 A; 2180 C; 1950 G; 1826 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0          Length:         8045
Score:                  7397.00    Matches:        1402
Percent Similarity:     100.00%    Conservative:   0
Best Local Similarity:  100.00%    Mismatches:     0
Query Match:            100.00%    Indels:         0
DB:                     3          Gaps:           0

US-09-913-878A-2 (1-1402) x AAA65171 (1-8045)

Qy          1 MetAsnProIleThrProArgLysArgAsnSerProValGluGluIleIleAsnArgLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2447 ATGAACCCTATTACTCCTAGGAAGAGGAATAGCCCCGTCGAGGAAATCATAAACCGGCTC 2506

Qy         21 AsnAsnAspTyrAsnLeuGlyLeuGlnCysValAlaAspThrThrLeuThrProHisArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2507 AATAACGACTACAACCTGGGCCTCCAGTGTGTCGCAGACACAACTCTCACCCCCCACCGC 2566

Qy         41 ArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHisAspLysIleTyrArgAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2567 CGGAAGGAGCTGGCCGAGAGTGACGAGGATTTCGGTCGCCATGACAAGATCTACAGAGCC 2626

Qy         61 LeuAsnPheLeuTyrTrpArgLysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2627 CTGAACTTTCTCTACTGGCGGAAGGATGACTCCCTGAACCAGGCAGAAGCCAACTTCTTC 2686

Qy         81 IleGluAlaLysAlaAlaSerSerAsnTrpValProLysAlaHisAlaAspProAspThr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2687 ATCGAGGCCAAAGCTGCGAGCTCGAACTGGGTGCCCAAAGCCCACGCCGACCCTGACACG 2746

Qy        101 LeuProTrpSerLysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGln 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2747 CTTCCGTGGTCCAAGGAACCTCCCCGCGCCGCTACTGCCGGCCAACAATGGGCATTGCAG 2806

Qy        121 ThrValLeuLeuGluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2807 ACTGTGTTGCTCGAGGTGCTTAATAGGTTTATGCCACCTCCCAATAACACACCAGGTCGA 2866

Qy        141 ThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThr 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2867 ACGTTTGGCAGAACTCTAAGCGGCCCAAGTGGCCTGAGCCGCCCAACCTCTACCAACACC 2926

Qy        161 LysArgLysAspGluProAlaAsnValThrPheAlaAspProProLysArgSerLeuThr 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2927 AAACGCAAGGATGAGCCCGCCAATGTCACTTTCGCTGATCCGCCCAAACGCTCGTTGACT 2986

Qy        181 ArgSerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2987 CGCTCTGCCACAGGTCCTCCTATTCACGGCGCGGCGATACCCCTAAAGTTCCCCGATCCA 3046

Qy        201 ValAsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLys 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3047 GTGAATACCGGTTCCAAACGACCATCTCTCGAGAGTGAGAATCTCAATCAGTGCACCAAG 3106

Qy        221 ArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValProIle 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3107 CGGGCCAAGGGCAAGCTGTCTGATAATGTTGCCGCTGCCGCCGCCCCGCCCGTGCCTATT 3166

Qy        241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3167 GCGAGCGCTTTGGACAAGGTACCGACTCGAAGGCATGCCAATACGAGAGATCCCACGGCG 3226

Qy        261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3227 ACAGGTCATAGACGAGCGGACCAGGTGGATTCCTTTGATACATCTCAAGGCACTTCCTAT 3286

Qy        281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3287 GGTTCGAGTGTCTTCAGCGCTTGCCGTCACAATCAGAGCACTACCCAGAGTAGTTTTGAG 3346

Qy        301 AlaProProSerGlnProArgGluLysArgProValAspAlaThrValPheGluAlaGly 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3347 GCTCCTCCTTCACAGCCCAGAGAGAAGCGGCCTGTGGATGCCACGGTCTTTGAGGCTGGA 3406

Qy        321 HisLeuIleGluSerProSerLysGlyArgThrThrLysSerHisIleAspAsnGlnPro 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3407 CACTTGATTGAGTCTCCTAGCAAAGGAAGAACAACCAAGTCCCACATAGATAACCAGCCC 3466

Qy        341 LeuSerSerSerSerGlnGlyGluThrSerPheSerThrTyrTyrGluSerPheProSer 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3467 CTTTCATCGTCTTCCCAGGGTGAAACTTCGTTCAGCACTTACTATGAGTCGTTTCCAAGT 3526

Qy        361 SerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLeuAlaArgSerGlu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3527 TCCGGCGGCGAGGGCGCAATTCCCGAGCCGAGTCGCTCAAATGGACTGGCTCGGAGCGAA 3586

Qy        381 GluSerAlaArgSerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsn 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3587 GAAAGCGCTCGATCTCAGGTTCAAGTTCATGCCCCGGTGGTTGCAGCTCGGCTGAGAAAT 3646

Qy        401 IleTrpProLysPheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluVal 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3647 ATTTGGCCGAAATTTCCCAAATGGCTACACGAAGCTCCTCTCGCTGTTGCATGGGAAGTT 3706

Qy        421 ThrArgLeuPheMetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyr 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3707 ACCAGACTCTTTATGCACTGCAAAGTAGACTTGGAAGACGAGAGCCTGGGCCTAAAGTAC 3766

Qy        441 AspProSerTrpSerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeu 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3767 GACCCTTCCTGGTCTACCGCGCGCGATGTCACAGATATCTGGAAGACTCTCTACCGGCTT 3826

Qy        461 AspAlaPheArgGlyLysProPheProGluLysProProAsnAspValPheValThrAla 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3827 GATGCTTTCCGTGGTAAACCCTTTCCAGAAAAGCCGCCCAACGACGTGTTCGTGACCGCA 3886

Qy        481 MetThrGlyAsnPheGluSerLysGlySerAlaValValLeuSerAlaValLeuAspTyr 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3887 ATGACGGGCAACTTTGAGAGCAAAGGTAGTGCCGTTGTTCTCTCTGCTGTTCTAGACTAC 3946

Qy        501 AsnProAspAsnSerProThrAlaProLeuTyrLeuValLysLeuLysProLeuMetPhe 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3947 AATCCGGACAACTCGCCTACTGCGCCCCTTTACCTTGTGAAGCTGAAGCCGCTCATGTTC 4006

Qy        521 GluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIleLeuIle 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4007 GAGCAGGGCTGTCGACTCACCCGTCGGTTCGGTCCTGATAGGTTTTTCGAGATCCTTATA 4066

Qy        541 ProSerProThrSerThrSerProSerValProProValValSerLysGlnProGlyAla 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4067 CCGTCGCCTACGAGCACCAGCCCAAGTGTACCGCCGGTGGTCAGCAAACAACCAGGTGCG 4126

Qy        561 ValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrp 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4127 GTCGAAGAAGTCATCCAGTGGCTCACGATGGGGCAACATTCTCTTGTAGGCCGCCAATGG 4186

Qy        581 ArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArg 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   4187 CGCGCTTTCTTCGCCAAAGATGCCGGATACAGGAAACCTCTCAGGGAGTTCCAGCTCCGC 4246

Qy    601 AlaGluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIle 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4247 GCCGAGGACCCGAAACCCATCATCAAGGAGAGAGTCCACTTCTTTGCCGAGACCGGCATT 4306

Qy    621 ThrPheArgProAspValPheLysThrArgSerValValProAlaGluGluProValGlu 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4307 ACGTTCCGACCTGATGTGTTCAAGACGAGATCTGTCGTTCCGGCAGAGGAACCTGTAGAG 4366

Qy    641 GlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAspAsnAsn 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4367 CAACGGACCGAGTTCAAAGTTAGTCAAATGCTGGACTGGCTCCTGCAACTCGACAACAAC 4426

Qy    661 ThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTyr 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4427 ACTTGGCAGCCGCACCTCAAGTTGTTCTCCCGTATCCAGCTCGGTCTGAGTAAGACATAT 4486

Qy    681 AlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAspLeuLeuSerPro 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4487 GCCATTATGACATTGGAGCCTCACCAGATCAGACACCACAAGACCGATCTTCTTTCACCT 4546

Qy    701 SerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSerValAlaLys 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4547 TCAGGCACTGGCGAAGTGATGAATGACGGTGTAGGCCGCATGTCGCGAAGCGTGGCCAAG 4606

Qy    721 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGly 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4607 AGGATACGCGATGTTCTCGGTTTGGGTGATGTGCCCTCTGCTGTGCAAGGGCGGTTTGGT 4666

Qy    741 SerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGlu 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4667 TCGGCCAAGGGAATGTGGGTTATCGACGTTGACGACACAGGCGATGAGGATTGGATCGAG 4726

Qy    761 ThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4727 ACATACCCGTCCCAGCGCAAGTGGGAATGCGACTTCGTTGATAAACATCAACGTACCCTC 4786

Qy    781 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuPro 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4787 GAAGTCCGGAGCGTGGCTTCTGAACTGAAGTCAGCTGGTCTCAACCTACAGCTGTTACCT 4846

Qy    801 ValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeu 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4847 GTCCTGGAAGATAGAGCCAGGGACAAGGTGAAGATGCGCCAGGCAATCGGTGACCGTCTT 4906

Qy    821 IleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4907 ATCAACGATTTGCAACGACAGTTCAGCGAGCAAAAGCATGCTTTGAATCGCCCAGTGGAA 4966

Qy    841 PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArg 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4967 TTTCGCCAATGGGTTTACGAGAGTTATTCCAGTCGCGCAACTCGAGTCAGCCACGGCCGT 5026

Qy    861 ValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsn 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5027 GTGCCTTTTCTTGCTGGGCTACCTGACAGTCAAGAGGAGACACTGAACTTCTTGATGAAC 5086

Qy    881 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5087 AGTGGGTTCGATCCCAAGAAGCAAAAGTACTTGCAAGACATCGCCTGGGATCTTCAAAAG 5146

Qy    901 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5147 CGGAAATGTGACACGTTGAAGTCCAAGCTGAACATCCGTGTCGGTCGATCAGCATACATT 5206

Qy    921 TyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5207 TACATGATTGCCGATTTCTGGGGTGTGCTTGAGGAAAATGAGGTTCATGTCGGATTCTCC 5266

Qy    941 SerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5267 TCAAAGTTCAGAGACGAGGAGGAGTCTTTTACACTCCTATCGGACTGTGATGTCCTCGTG 5326

Qy    961 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5327 GCGCGATCCCCAGCCCATTTCCCTAGTGATATCCAACGGGTTCGAGCAGTCTTCAAGCCA 5386

Qy    981 GluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAla 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5387 GAGCTCCACAGTCTCAAGGATGTAATCATCTTCTCTACTAAAGGAGATGTACCGCTTGCT 5446

Qy   1001 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5447 AAGAAGCTATCTGGTGGAGACTACGACGGCGATATGGCCTGGGTCTGCTGGGATCCGGAG 5506

Qy   1021 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5507 ATCGTCGATGGTTTCGTCAATGCGGAAATGCCTCTGGAGCCCGACCTGTCTAGGTACCTA 5566

Qy   1041 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5567 AAGAAGGACAAAACGACTTTCAAACAACTTATGGCCTCACACGGCACGGGCTCAGCGGCC 5626

Qy   1061 LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5627 AAAGAGCAGACTACATACGATATGATCCAGAAGAGCTTCCATTTCGCCCTGCAGCCCAAC 5686

Qy   1081 PheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSer 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5687 TTCTTGGGCATGTGCACTAACTACAAAGAAAGGCTCTGTTACATCAACAATAGTGTGTCT 5746

Qy   1101 AsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGln 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5747 AACAAGCCGGCCATCATTCTTAGTTCACTGGTGGGAAACCTCGTCGATCAGAGCAAGCAA 5806

Qy   1121 GlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAla 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5807 GGTATTGTCTTTAACGAAGCAAGCTGGGCTCAATTGCGTAGGGAACTGCTTGGCGGTGCA 5866

Qy   1141 LeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThr 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5867 TTGTCCCTTCCTGACCCAATGTACAAGAGCGACAGTTGGCTCGGGCGCGGAGAGCCTACC 5926

Qy   1161 HisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGlu 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5927 CACATTATTGACTACCTGAAATTCTCCATCGCCAGGCCTGCGATTGACAAGGAACTGGAA 5986

Qy   1181 AlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAsp 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5987 GCCTTCCACAATGCCATGAAAGCGGCCAAGGATACAGAAGACGGCGCTCACTTTTGGGAT 6046

Qy   1201 ProAspLeuAlaSerTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSer 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6047 CCGGATCTCGCTTCCTACTACACGTTCTTCAAGGAGATTAGCGACAAGTCGCGATCGTCC 6106

Qy   1221 AlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArg 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6107 GCACTGCTATTCACGACTCTGAAGAACCGTATCGGCGAAGTCGAGAAAGAATATGGCAGG 6166

Qy   1241 LeuValLysAsnLysGluMetArgAspSerLysAspProTyrProValArgValAsnGln 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6167 TTGGTCAAAAACAAGGAGATGAGAGACAGCAAGGACCCCTACCCTGTCCGCGTCAACCAG 6226

Qy   1261 ValTyrGluLysTrpCysAlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyr 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6227 GTTTATGAAAAATGGTGCGCCATCACGCCTGAGGCGATGGACAAATCCGGAGCAAATTAT 6286

Qy   1281 AspSerLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThr 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6287 GATTCTAAGGTGATCAGGTTGCTGGAGCTGTCCTTCCTCGCGGACCGTGAGATGAATACA 6346

Qy   1301 TrpAlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPhe 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6347 TGGGCATTGCTGAGGGCTAGCACGGCTTTCAAGCTGTACTACCACAAGAGCCCCAAGTTC 6406
```

```
Qy       1321 ValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArgPro 1340
Db       6407 GTGTGGCAGATGGCGGGCAGACAGCTCGCGTACATTAAGGCGCAGATGACGAGCAGACCC 6466

Qy       1341 GlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLys 1360
Db       6467 GGTGAAGGCGCCCCGGCGTTGATGACCGCGTTCATGTATGCGGGCTTGATGCCGGATAAG 6526

Qy       1361 LysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAspPro 1380
Db       6527 AAGTTTACGAAGCAGTATGTGGCCAGGCTGGAGGGCGATGGATCGGAGTACCCTGATCCG 6586

Qy       1381 GluValTyrGluValLeuGlyAspAspAspPheAspGlyIleGlyPheThrGlyAsnGly 1400
Db       6587 GAGGTCTATGAAGTGCTGGGCGATGATGATTTTGATGGAATTGGTTTCACAGGGAATGGC 6646

Qy       1401 AspTyr 1402
Db       6647 GATTAT 6652
```

```
RESULT 2
ABZ54843
ID   ABZ54843 standard; cDNA; 615 BP.
XX
AC   ABZ54843;
XX
DT   28-MAR-2003  (first entry)
XX
DE   Aspergillus oryzae polynucleotide SEQ ID NO 3956.
XX
KW   Aspergillus oryzae; fermentation; fungus; industrial; EST;
KW   expressed sequence tag; gene; ss.
XX
OS   Aspergillus oryzae.
XX
PN   WO200279476-A1.
XX
PD   10-OCT-2002.
XX
PF   22-MAR-2002; 2002WO-IB000890.
XX
PR   30-MAR-2001; 2001JP-00098371.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (NARE-) NAT RES INST BREWING.
PA   (NORQ ) NAT FOOD RES INST MIN AGRIC.
XX
PI   Machida M,  Akita O,  Kashiwagi Y,  Kitamoto K,  Horiuchi H;
PI   Takeuchi M,  Kobayashi T,  Kitamoto N,  Gomi K,  Abe K;
XX
DR   WPI; 2003-046817/04.
XX
PT   Detection of expression of specific Aspergillus genes for monitoring the
PT   fermentation and growth conditions of the fungus, using DNA probes.
XX
PS   Claim 1; SEQ ID NO 3956; 48pp + Sequence Listing; Japanese.
XX
CC   The invention relates to a polynucleotide having any of 6006 specific
CC   sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
CC   specific culture conditions including one or more of eutrophic,
CC   oligotrophic, solid, early germination, alkaline, high temperature, low
CC   temperature or maltose culture or polynucleotides stringently hybridising
CC   to these sequences. The polynucleotides are useful for monitoring the
CC   progress of fermentation and the growth conditions of a fungus,
CC   especially of Aspergillus oryzae which is widely used in industrial
CC   fermentation. Also monitoring for fungal contamination. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 615 BP; 151 A; 148 C; 162 G; 154 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.41e-12     Length:         615
Score:                  341.00       Matches:        89
Percent Similarity:     54.75%       Conservative:   32
Best Local Similarity:  40.27%       Mismatches:     80
Query Match:            4.61%        Indels:         21
DB:                     7            Gaps:           6

US-09-913-878A-2 (1-1402) x ABZ54843 (1-615)
```

```
Qy        935 ValHisValGlyPheSerSerLysPheArg-AspGluGluGluSerPhe------ThrLe 952
Db          1 GTTCACTTTGGCTTCTCCAGTAACTGGCTGTGATCCTGAAGGCCAATTCGAAGACAATCT 60

Qy        952 uLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGl 972
Db         61 ACTCGACGGTATGGACGTTCTTGTAGCTCGGCTCCCCGCACATGTGCCGTCTGATATCCA 120

Qy        972 nArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSe 992
Db        121 GCGACGAAGAGCTGTGTGGAAGCCAGAGCTCCGGCATTTCAAAGATGTTATCGTTTTCCC 180

Qy        992 rThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMe 1012
Db        181 TACTGTTGGGACAACGCCACTGGCGCACATGTTATCCGGTGGTGATTACGACGGTGATAC 240

Qy       1012 tAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet---Pr 1031
Db        241 AACGTGGGTTTGCTGGGACCAAAATATTGTCCAGAAGTTTCGTAATTCCGACCTATCCAC 300

Qy       1031 oLeuGluProAspLeuSerArgTyr---LeuLysLysAspLysThrThrPheLysGlnLe 1050
Db        301 CATGGAATACCCTGCGGAGCACTTTGGTCTTGAGGAACATAACGTGCCAATGAAAGATAT 360

Qy       1050 uMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGl 1070
Db        361 T----------------------------------GACTCATGGGACGAGTTTCTTCA 384

Qy       1070 nLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGl 1090
Db        385 AAGCACTTTTACGTTCAATTTAACCATGTCTAATCTGGGAAGGTGTACTGTCGAGCACGA 444

Qy       1090 uArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLe 1110
Db        445 GAAGATATCCTAC---GACGAATCGATCGACTCGCCGAATGCCAAAGAGCTCGCCTGTCT 501

Qy       1110 uValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAl 1130
Db        502 CCTCAGTCATTTGGTGGACGGTCGTAAAGGAGGGGTCCGCTTATCAGACAG-GCGTGGCG 560

Qy       1130 aGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSe 1150
Db        561 AGCATATCGCAAAAGGGTTAGCCCTAGGGCACGGGATGTG------CCTGCGTACAAGAA 614

Qy       1150 r 1150
Db        615 C 615
```

```
RESULT 3
AAH77704
ID   AAH77704 standard; cDNA; 3591 BP.
XX
AC   AAH77704;
XX
DT   13-NOV-2001  (first entry)
XX
DE   Nucleotide sequence of a plant SGS2 polypeptide.
XX
KW   SGS2; RNA-dependent RNA polymerase; transgene silencing;
KW   transgene stability; crop plant; viral resistance; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             1. .3591
FT                   /*tag=  a
FT                   /product= "SGS2"
XX
PN   FR2804128-A1.
XX
PD   27-JUL-2001.
XX
PF   26-JAN-2000; 2000FR-00001007.
XX
PR   26-JAN-2000; 2000FR-00001007.
XX
PA   (RHOB-) RHOBIO.
PA   (INRG ) INST NAT RECH AGRONOMIQUE.
XX
PI   Beclin C,  Elmayan T,  Mourrain P,  Vaucheret H;
XX
DR   WPI; 2001-543303/61.
DR   P-PSDB; AAG67235.
XX
PT   A new plant SGS2 gene involved in encoding an RNA-dependent RNA
PT   polymerase and in transgene silencing, increases transgene stability and
PT   expression in transgenic plants when it is inactivated.
XX
PS   Claim 4; Page 32-37; 46pp; French.
XX
CC   The present sequence encodes a plant SGS2 polypeptide. SGS2 is an RNA-
CC   dependent RNA polymerase and is involved in transgene silencing.
CC   Inactivation of SGS2 is used to increase transgene stability and
CC   expression in plants, particularly crop plants, especially maize, corn,
CC   barley, sorghum, soya, sugar cane, beet, tobacco or cotton plants.
CC   Overexpression of SGS2 can be used to increase resistance to viral
CC   infection in plants
XX
SQ   Sequence 3591 BP; 1022 A; 714 C; 858 G; 997 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.21e-10      Length:        3591
Score:                  328.00        Matches:       197
Percent Similarity:     37.17%        Conservative:  126
Best Local Similarity:  22.67%        Mismatches:    312
Query Match:            4.43%         Indels:        234
DB:                     4             Gaps:          42

US-09-913-878A-2 (1-1402) x AAH77704 (1-3591)

Qy        535 PhePheGluIleLeuIleProSerProThrSerThrSerProSerValProProValVal 554
                |||         ::::::           ||||||:::|||   |||              |||
Db       1396 TTTGTTGCTCCTATTGTGAAGGATTTGACATCAAGTTCTTTCTCCCAGAAGACCTACGTT 1455

Qy        555 SerLysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSer 574
                 |||:::     ::::::                     ||||||   |||
Db       1456 TTTAAAAGAGTGAAGAGCATA------------------TTAACCGATGGG---TTTAAA 1494

Qy        575 LeuValGlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeu 594
              |||   ||||||::::::   :::|||   |||      |||
Db       1495 CTATGTGGTAGAAAATAC---AGTTTTCTAGCATTCTCAGCCAAT--------------- 1536

Qy        595 ArgGluPheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhe 614
                       |||||||||                                       |||
Db       1537 ---------CAACTGAGA------------------------GACCGCTCTGCATGGTTC 1563

Qy        615 PheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValPro 634
              |||||||||   |||                     |||||||||
Db       1564 TTTGCTGAAGACGGG-----------------------AAAACACGT------------- 1587

Qy        635 AlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeu 654
                                                      ||||||   :::   |||:::
Db       1588 -------------------------------------GTGTCAGATATAAAAACATGGATG 1611

Qy        655 LeuGlnLeu---AspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGln 673
                 :::         |||   |||                  |||      :::|||:::
Db       1612 GGGAAGTTCAAAGACAAGAATGTG------------GCAAAATGTGCTGCTAGGATGGGC 1659

Qy        674 LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHis 693
              |||         |||   ||||||||   :::   :::   |||||||::::::
Db       1660 CTGTGCTTCTCCTCCACATATGCCACTGTAGATGTCATGCCTCACGAGGTT--------- 1710

Qy        694 LysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArg 713
                 |||::::::                  |||   |||   :::||||||:::|||
Db       1711 GACACTGAGGTTCCAGATATTGAGAGAAATGGGTATGTTTTCTCTGACGGAATTGGTACA 1770

Qy        714 MetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal------ 731
              ::::::      :::|||   :::   :::   |||   |||   ||||||
Db       1771 ATCACACCTGACCTCGCTGACGAAGTAATGGAGAAACTTAAGTTG---GATGTGCACTAC 1827

Qy        732 ---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspVal 750
                 |||   |||   |||   |||:::         ||||||:::
Db       1828 AGCCCTTGTGCTTATCAGATACGTTACGCAGGTTTCAAAGGGGTT--------------- 1872

Qy        751 AspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln--------------- 765
                                    :::      :::|||||||:::
Db       1873 ------------------------GTTGCTCGTTGGCCATCAAAAAGTGATGGAATCAGG 1908

Qy        766 ---------ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluVal 782
                       |||            |||      ||||||         ||||||:::
Db       1909 CTAGCCCTTCGAGACAGTATGAAGAAGTTCTTTTCCAAACAT---ACGATCTTGGAGATC 1965

Qy        783 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 802
                 |||      :::   :::            ||||||   |||::::::   :::|||
Db       1966 TGTTCC---TGGACGAGGTTTCAACCTGGGTTCTTAAATCGGCAGATAATTACCCTTCTA 2022

Qy        803 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 822
                                                   :::   |||   :::
Db       2023 TCC--------------------------GTACTAGGTGTTCCGGATGAAATATTCTGG 2055

Qy        823 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArg 842
               |||:::|||                     :::   |||||||||     ::::::
Db       2056 GATATGCAG---------GAATCCATGCTCTATAAACTGAACCGCATCCTTGATGATACA 2106

Qy        843 GlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValPro 862
                    :::|||      ::::::
Db       2107 GATGTGGCATTTGAAGTTCTCACGGCATCATGTGCT------------------------ 2142

Qy        863 PheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGly 882
                                    |||      |||         ::::::::::::|||
Db       2143 --------------------GAACAGGGAAACACTGCAGCTATCATGCTTAGTGCAGGT 2181

Qy        883 PheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLys 902
              |||   ||||||   :::   :::|||:::   :::         ::::::      :::
Db       2182 TTCAAACCAAAAACCGAGCCGCATCTACGCGGGATGTTGTCTTCAGTCAGAATTGCACAA 2241

Qy        903 CysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMet 922
                          |||:::   |||      |||   |||      ::::::
Db       2242 CTCTGGGGTCTCAGAGAAAAATCTCGTATTTTTGTTACTTCAGGAAGGTGGCTAATGGGT 2301

Qy        923 IleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer------ 940
                 |||      |||:::||||||         :::      :::         |||
Db       2302 TGCCTAGACGAAGCAGGGATACTTGAACATGGCCAATGCTTTATTCAAGTCTCTAAACCG 2361

Qy        941 ------------------------------SerLysPheArgAspGluGluGluSerPhe 950
                                            |||:::|||::::::   :::
Db       2362 TCTATAGAAAATTGTTTCTCCAAACATGGGTCTCGTTTTAAGGAGACAAAGACAGATCTG 2421

Qy        951 ThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAsp 970
              ::::::      |||   :::|||::::::|||         |||   |||
Db       2422 GAAGTAGTTAAAGGCTAT---GTAGCCATTGCTAAGAATCCTTGTCTTCACCCAGGGGAT 2478

Qy        971 IleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIle 990
              ::::::   :::   ||||||      |||:::||||||   :::   |||   :::|||
Db       2479 GTAAGGATTTTAGAAGCTGTTGATGTACCCCAGCTGCATCACATGTATGACTGCCTTATT 2538

Qy        991 PheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGly 1010
              |||      |||||||||   |||         :::   ||||||   |||   ||||||
Db       2539 TTCCCTCAGAAAGGTGATAGGCCGCATACAAACGAAGCTTCTGGCAGTGACCTTGACGGG 2598
```

```
Qy       1011 AspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet 1030
              |||:::   :::|||  |||||   :::::::::
Db       2599 GACCTGTACTTTGTGGCTTGGGATCAGAAACTCATC----------------------- 2634

Qy       1031 ProLeuGluProAspLeuSerArgTyr------LeuLysLysAspLysThrThrPheLys 1048
                       |||:::        |||      :::      |||            |||
Db       2635 ------CCTCCCAACAGGAAAAGCTATCCGGCCATGCATTATGATGCAGCTGAAGAGAAG 2688

Qy       1049 GlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMet 1068
                 |||                 |||   |||         |||         |||
Db       2689 AGTTTA-----------------GGCCGTGCTGTCAACCACCAGGACATAATCGATTTC 2730

Qy       1069 IleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsn--- 1087
                    ::::::           |||            |||||         |||
Db       2731 TTTGCAAGAAAC-----------TTGGCGAATGAGCAGTTGGGCACAATTTGCAATGCA 2778

Qy       1088 ---------TyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAla 1104
                       :::   :::|||      |||               ::::::
Db       2779 CACGTCGTTCATGCTGATAGAAGTGAGTAT---------GGAGCCATGGACGAAGAATGT 2829

Qy       1105 IleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPhe 1124
              ::::::|||:::   |||          ||||||     |||   |||    :::
Db       2830 TTGCTACTGGCAGAACTAGCTGCCACTGCAGTTGATTTCCCAAAGACAGGGAAAATTGTG 2889

Qy       1125 AsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuPro 1144
              :::                     |||:::   :::|||         :::
Db       2890 TCAATGCCC------TTCCACCTAAAACCAAAACTCTACCCAGATTTCATGGAAAAAGAA 2943

Qy       1145 Asp---ProMetTyrLysSerAspSerTrpLeuGlyArg-------------------- 1156
              |||            |||||||||:::      ||||||||
Db       2944 GACTACCAAACTTACAAGTCGAACAAAATCTTGGGTCGGCTTTACAGACGGGTAAAAGAG 3003

Qy       1157 ---------------------------------------GlyGluProThrHisIle--- 1162
                                                     ::::|||:::   |||
Db       3004 GTTTATGATGAAGATGCAGAAGCTTCCTCAGAAGAAAGCACAGACCCAAGTGCCATCCCT 3063

Qy       1163 ------------IleAspTyrLeuLysPheSerIleAlaArgPro--------------- 1173
                          ::::::            |||         |||
Db       3064 TATGACGCTGTTCTTGAAATACCGGGATTTGAAGATTTGATCCCTGAGGCATGGGGTCAC 3123

Qy       1174 ------AlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAsp 1191
                          |||   :::|||                  |||      |||:::
Db       3124 AAATGTTTGTACGACGGGCAACTCATTGGTCTTCTTGGGCAATACAAGGTGCAGAAAGAG 3183

Qy       1192 ThrGlu---AspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePhe 1210
                 |||               |||   |||                  |||||
Db       3184 GAAGAGATTGTGACGGGTCACATCTGGTCC---------ATGCCCAAATACACAAGCAAG 3234

Qy       1211 LysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArg 1230
              |||:::         |||   |||                        ||||||:::
Db       3235 AAACAAGGCGAACTGAAAGAAAGA-----------------------CTGAAGCATTCT 3270

Qy       1231 IleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSer 1250
                       ::::::||||||:::   |||   |||         :::   :::   |||
Db       3271 TATAATTCCCTTAAAAAGGAGTTC---AGGAAAGTATTTGAGGAAACAATCCCTGACCAT 3327

Qy       1251 LysAspProTyrProValArgValAsnGlnValTyrGluLys------------TrpCys 1266
              ::::::                        |||   :::||||||||            |||
Db       3328 GAAAATCTCAGCGAGGAGGAAAAAAACATCTTGTATGAGAAGAAGGCTTCAGCTTGGTAT 3387

Qy       1267 AlaIleThr------ProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysVal 1284
                 :::|||         ||||||   :::   ||||||
Db       3388 CATGTAACTTACCATCCCGAGTGGGTGAAGAAGTCT---------------------- 3423

Qy       1285 IleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrpAlaLeuLeu 1304
                       |||||||||               |||   :::   ::::::   |||   :::
Db       3424 ---------TTAGAGCTGCAA---------GATCCAGATGAGTCCAGTCATGCGGCGATG 3465

Qy       1305 ArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPheValTrpGlnMet 1324
                 :::                                       |||   |||   :::
Db       3466 CTGAGT---------------------------------------TTTGCTTGG---ATT 3483

Qy       1325 AlaGlyArgGlnLeuAlaTyrIleLys 1333
```

```
                  |||              ||||||   ||||||
Db       3484 GCAGCTGATTATCTTGCAAGAATCAAA 3510

RESULT 4
AAC63740
ID   AAC63740 standard; cDNA; 3901 BP.
XX
AC   AAC63740;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Maize RNA-directed RNA polymerase FIS sequence, SEQ ID NO: 7.
XX
KW   Maize; plant; RNA-directed RNA polymerase; FIS; gene mapping;
KW   gene marker; plant virus resistance; plant breeding; ss.
XX
OS   Zea mays.
XX
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT,  Orozco EM,  Weng Z,  Sakai H,  Cahoon RE,  Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   P-PSDB; AAB28531.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 2; Page 40-41; 62pp; English.
XX
CC   The present sequence is one of a number of cDNA molecules which encode
CC   plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC   probes for genetically and physically mapping genes, and as markers for
CC   traits linked to those genes. They are useful for controlling gene
CC   expression and provide mechanisms to engineer plant virus resistance.
CC   They are also useful for plant breeding to develop lines with desired
CC   phenotypes
XX
SQ   Sequence 3901 BP; 928 A; 975 C; 1058 G; 935 T; 0 U; 5 Other;

Alignment Scores:
Pred. No.:                1.15e-09      Length:          3901
Score:                    319.50        Matches:         290
Percent Similarity:       33.81%        Conservative:    185
Best Local Similarity:    20.64%        Mismatches:      528
Query Match:              4.32%         Indels:          405
DB:                       3             Gaps:            62

US-09-913-878A-2 (1-1402) x AAC63740 (1-3901)

Qy         92 ProLysAlaHisAlaAspProAspThrLeuProTrp------------SerLysGluPro 107
              |||         |||      |||      ||||||            :::::::::|||
Db         24 CCCAACGGATCCGCGGCAACCGACCAC---CCATGGGATCGCTCCGGGGCGCGGCAGCCT 80

Qy        108 ProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeuGluValLeu 127
              |||               |||         |||      :::   :::   |||      |||
Db         81 CCTCCGCGGCGCCGCGCGCGGGCGACCTGGTGACCACGCAGGTTAGCCTTGGTGGATTTG 140

Qy        128 AsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArgThrLeuSer 147
                       |||||||||                              |||::::::|||
Db        141 ---------ATGCCACCGTCAAGG----------------------CGCTCGATCTCG 167
```

```
Qy    148 GlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAspGluProAla 167
Db    168 --------------------CCGACTTCCTCGAGTTGAATGCGG--------------- 191
Qy    168 AsnValThrPheAlaAspProProLysArgSerLeuThrArgSerAlaThrGlyProPro 187
Db    192 ------------------------GCTCGGTCTGGCGCTGCCGCCTTAAGACCTCCTGGA 227
Qy    188 IleHisGlyAlaAlaIleProLeuLysPheProAspProValAsnThrGlySerLysArg 207
Db    228 CTCCGCCGGACGCCTATCCCGACTTCCTTCTCCCCACCG------------TCACCTCCG 275
Qy    208 ProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArgAlaLysGlyLysLeuSer 227
Db    276 CCGCCGCGCCGCCGCCACAGTACGATCGCGTGCCTCCGCACGCCTTCGTCCACTTTG--- 332
Qy    228 AspAsnValAlaAlaAlaAlaAlaProProValProIleAlaSerAlaLeuAspLysVal 247
Db    333 ---CGCGCCCGGAGGGCGCGCGCGCCGCCG----------------------------- 359
Qy    248 ProThrArgArgHisAlaAsnThrArgAspProThrAlaThrGlyHisArgArgAlaAsp 267
Db    360 -----------------------------CCGACGCAAGCGGGC------CGATCCGAG 383
Qy    268 GlnValAspSerPheAspThrSerGlnGlyThrSerTyrGlySerSerValPheSerAla 287
Db    384 CTCATCCTYTCC---------GGCAAACCCCTGCGCGCCGCCTCCGCACAGGACAGCTCC 434
Qy    288 CysArg-HisAsnGlnSerThrThrGlnSerSerPheGluAlaProProSerGlnProAr 307
Db    435 CTTCGGGCATCCCGCCGCCGTAAGTGTCTCGCCATTCCGCTTCCCTGGCTCGCGC----- 489
Qy    307 gGluLysArgProValAspAlaThrValPheGluAlaGlyHisLeuIleGluSerProSe 327
Db    490 ---------------------------CTCGAGGTCGGG---------GNNTCTCCCGG 512
Qy    327 rLysGlyArgThrThrLysSerHisIleAspAsnGlnProLeuSerSerSerSerGlnGl 347
Db    513 CCTCGGACGCCTTCATCGCCGCCTGGCGCGGCCCCGGCCTCTGGGCTCGAGTTCTCCGTC 572
Qy    347 yGluThrSerPheSerThrTyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIl 367
Db    573 GACCCGTTCGACGGGTCTTGCCNCTTCATCTTCGCCCGCGACA----------------- 615
Qy    367 eProGluProSerArgSerAsnGlyLeuAlaArgSerGluGluSerAlaArgSerGlnVa 387
Db    616 -------CCGCTTTCAAAGTCCGGGAG-TTCCGCGAGTCTGTGGTGATGCGCTGCGACGT 667
Qy    387 lGlnValHisAlaPro---------ValValAlaAlaArgLeuArgAsnIleTrpProLy 404
Db    668 CAAGCTCCAGTTCCCCGTCCGCGATGTTGCGGAAGTCAGGGTGTTCCGGCTCGACTGCTC 727
Qy    404 sPheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPh 424
Db    728 GCTGCTGATCCGGCTATCGGCCGCACCGCTG---GTCTGTTACCGCACGGCGGATGACGA 784
Qy    424 eMetHisCysLysVal------AspLeuGluAspGluSerLeuGlyLeuLysTyrAspPr 442
Db    785 CATCCACGTGTCCGTGCCGTTCGACCTGCTCGACGAC------------------GATGA 826
Qy    442 oSerTrpSerThrAlaArgAspValThrAsp---------------------------- 452
Db    827 CCCGTGGATACGGACCACGGACATCACTCCAAGTGGTGCGATTGGGCGGTGCGGTGCGTA 886
Qy    453 ----------------------IleTrpLysThrLeuTyrArg----------------- 459
Db    887 TAGAATCACATTCTCGCCGCGGTTCTGGCCAAAGATGGAACGCGCGCTGACGTACATGAG 946
Qy    460 ------------------LeuAsp------------------AlaPheArgGlyLysPr 467
Db    947 GGATAGGAGGGTGGCGATCCTTGATTGCGTTGSAGGGTGGGGGCCAGGAGGGGGCTCAC 1006
Qy    467 oPheProGluLysProPro---------------AsnAspValPhe------------- 477

Db   1007 CGTGCGTGATGAGCCTGAGTTTGGGGAGCGGATGCAGGACCTGTTCTTCTGCGTGCAGCA 1066
Qy    478 ----------------------------------ValThrAlaMetThr----------- 482
Db   1067 CGCCGAGGGTCTCAAGTTTCCGGTGTTGTTCCTCGTGAATGCTCTGGTGCACAAGGGAGT 1126
Qy    483 ----------------------------------GlyAsnPheGluSerLysGlySerAl 491
Db   1127 AATAAGTCAACACCACCTCACGCCTGAATTCTTCGGTTTGCTCCAGAGGAAGGAGGATGA 1186
Qy    491 aValValLeuSerAlaValLeuAspTyrAsnProAspAsnSerProThr----------- 507
Db   1187 TGTGAATGTGGCTGCCTTGAGGGAATTTTGGGGGGACAAATTTCCAGTTTTGATGCATG 1246
Qy    508 ---------------------------AlaProLeuTyrLeuValLysLeuLysProLe 518
Db   1247 TGGGAGGCTGAAGAATCTGCAGGATAGGGTTGCCAGGTACCTGAAACATCTTCGCAACAA 1306
Qy    518 uMetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIl 538
Db   1307 GATTGGGGATGTCAATTCTGAGGTGAGGAGG----------------------------- 1337
Qy    538 eLeuIleProSerProThrSerThrSerProSerValProProValValSerLysGlnPr 558
Db   1338 -CTGGTAATCACGCCCACCAAG---GCTTATTGCATGCCACCAGAAGTGGAGCGCTCT-- 1391
Qy    558 oGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyAr 578
Db   1392 ----------AATCGCGTCATCCGG---------------CATTATAGTGAAGTCTCAGA 1426
Qy    578 gGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArg-------------------- 591
Db   1427 CCGGTTTCTGAGGGTTACTTTTATGGATGAGGGAATGCAGATGCTCAACAGTAATGTGCT 1486
Qy    592 ----------------LysProLeuArgGluPheGlnLeuArgAlaGluAspProLysPr 606
Db   1487 GAATTTCTCTGCTGCACAAATCGTCAAAGATTTGATGTCAAACTCGTTCCTGCATAAGAC 1546
Qy    606 oIleIleLysGluArgValHisPhePheAlaGluThrGly-------------------- 619
Db   1547 AACAGTGTACAAGCGTGTTAAAACGTTTTTGACAGAGGGATTCCACATGTGTGGCAGGAA 1606
Qy    620 ----------IleThrPheArgProAspValPheLysThrArgSerValValProAlaGl 636
Db   1607 GTACTCGTTTCTTGCATTCTCATCTAACCAGCTGAGGGACAGGTCA---------GCATG 1657
Qy    636 uGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGl 656
Db   1658 GTTCTTCGCAGAGGACAGAACGACA---ACAGTGGAAACCATTAGGAAATGGATGGGGCG 1714
Qy    656 nLeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLe 676
Db   1715 GTTCACAAGTAAGAATGTAGCAAAGCATGCC---------GCTCGGATGGGGCAGTGCTT 1765
Qy    676 uSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAs 696
Db   1766 CTCCTCTACATATGCTACGGTGGTGCTGCAGCCGCATGAGGTAAATGAGTGTCTTGATGA 1825
Qy    696 pLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerAr 716
Db   1826 AGTT---------GAACATAACGGGTACATTTTCTCTGATGGAATTGGCAAGATTACGTG 1876
Qy    716 gSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAsp---ValProSerAlaVa 735
Db   1877 CGACCTTGCACTCGAAGTTGCTCAGAAGCTGCAATTGACAGATAATCCCCCATCTGCTTA 1936
Qy    735 lGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspVal---AspAspThrGl 754
Db   1937 CCAGATTAGGTATGCAGGCTTCAAGGGT------GTTATATCTGTCTGGGAAGGAAAAAA 1990
Qy    754 yAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAspPheValAs 774
```

```
Db       1991 TGATGGGATACGACTTTCCCTGAGGCCGAGCATGCACAAGTTTGAG------------- 2036

Qy        774 pLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLe 794
                :::        ||||||||     |||       ::::::   :::          ||
Db       2037 -TCTAACCATACTGTGTTAGAGGTGGTCTCG---TGGACAAAGTTTCAGCCAGGATTCTT 2092

Qy        794 uAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGl 814
              ||||    |||::::::    :::|||
Db       2093 AAATCGTCAGATTATTACATTACTGTCC------------------------------- 2120

Qy        814 nAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAl 834
                                                                          ::
Db       2121 ----------------------------------------------------------TC 2122

Qy        834 aLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaTh 854
              :||||||   |||                          :::    |||:::      |||
Db       2123 CTTGAATGTCCCGGATGCTATCTTTGCTCAAATGCAGGAAGCCATGTTATCTAATCTCAA 2182

Qy        854 rArgVal---SerHisGlyArgValProPhe------LeuAlaGlyLeuProAspSerGl 871
                :::    |||          |||   |||                           ||
Db       2183 CAATATTTTGTCAGACTCTGATGTTGCTTTTGACATTGTAACCGCCTCTTGTGCTGAGCA 2242

Qy        871 nGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLe 891
              |        |||          :::::::::::|||      |||       :::    :::||
Db       2243 AGGAACCACTGCAGCACTGATGTTGAGTGCTGGCATTTCACCTGGAACTGAGCCACACCT 2302

Qy        891 uGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAs 911
              |:::    :::       ::::::       :::           |||         |||
Db       2303 GAAAGCAATGCTGTTAGCTATAAGGTCCTCACAGCTGCTAGGTCTTTTGGAGAAGACAAG 2362

Qy        911 nIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGl 931
               |||   |||   |||          ::::::          |||       |||:::||||
Db       2363 GATTTTTGTGCCCAAGGGAAGGTGGTTGATGGGCTGCCTTGATGAACTTGGGATCCTTGA 2422

Qy        931 uGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluSerPheTh 951
              |:::    :::    :::       ||||||
Db       2423 GCAAGGGCAGTGCTTTATCCGGGCATCATCT-------------------------CC 2455

Qy        951 rLeuLeuSerAspCys--------------------------------------------- 956
                    |||::::::|||
Db       2456 ATCACTCAATAATTGTCTGGTGAAGTATGGATCAAGATTGTCTGCAGCAAACACAAATGC 2515

Qy        957 ---------------AspValLeuValAlaArgSerProAlaHisPheProSerAspIl 971
                             ::::::::|||::::::|||          |||     |||::
Db       2516 AGAGACCATTCTGGGTACTATCGTAATGGCAAAGAATCCATGCCTTCATCCAGGGGATGT 2575

Qy        971 eGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePh 991
              ::::    :::    ||||||       ||||||||||||    |||    |||    ::::::||
Db       2576 CCGAATCCTTGAAGCTGTTGATGTGCCTGAACTGCATCACCTTGTTGATTGCTTGGTCTT 2635

Qy        991 eSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAs 1011
              |          ||||||:::    |||    |||    :::    ||||||    |||    ||||||||
Db       2636 CCCCAAGAAAGGTGAGAGGCCGCACGCGAATGAAGCATCTGGGAGTGATCTTGATGGGGA 2695

Qy       1011 pMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetPr 1031
              |:::    :::|||   ||||||       ::::::                 :::         ||
Db       2696 TCTATACTTCGTAACATGGGATGAAAACCTTATACCACCTGGTAAAAAGAGTTGGAACCC 2755

Qy       1031 oLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMe 1051
              |::::::              |||          :::             :::    ::
Db       2756 TATGGAC---------------TACTCCCCAGCTGAAGCAAAACAACTGCCACGCGCAGT 2800

Qy       1051 tAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLy 1071
              ::::    |||                                   |||::::::
Db       2801 ATCCCAACAT-----------------------------------GATATTGTT----- 2819

Qy       1071 sSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluAr 1091
                 |||                  ||||||       |||          |||::
Db       2820 -GGTTTC--------------------TTCTTGAAGAACATGGTAAAT------GAGAA 2851

Qy       1091 gLeuCysTyrIleAsnAsnSer--------------------------------ValSe 1100
              :|||        |||:::|||:::
Db       2852 ACTGGGTCCAATAAGCAATGCTCATGTTGTTCACGCTGATATGAGCGAGTATGGAGCAAT 2911

Qy       1100 rAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGl 1120
               ::::::          |||    |||:::    |||                |||||||    |||
Db       2912 GGATGAGAAGTGTATTCAGTTGGCAGAACTAGCAGCAACTGCTGTGGACTTCCCCAAGAC 2971

Qy       1120 nGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAl 1140
               |||    :::    :::              |||    ||||||    :::|||
Db       2972 AGGCAAAATTGTGTCAATGCCA------GCATCCCTTCGACCAAAATTATATCCTGACTT 3025

Qy       1140 aLeuSerLeuProAspProMet---TyrLysSerAspSerTrpLeuGlyArg-------- 1156
               :::          |||    :::    |||:::|||:::          |||||||||
Db       3026 CATGGGAAAGGAGGATGCTATCTCCTATAGATCAGAGAAGATCCTTGGAAGGCTTTATCG 3085

Qy       1156 ------------------------------------------------------------ 1156

Db       3086 GTCAATCCAAGAAGCCTCCAGCGATGATTTGGTTCCAGAAGAAACTTGCACATCTAACAA 3145

Qy       1157 ------------------GlyGluProThrHisIleIleAspTyrLeuLysPheSer-- 1169
                                |||                     |||:::|||    :::
Db       3146 TCTGCCTTATGATGCAGATATGGAAGTTGCTGGTGCAGCTGATTTTCTCTCGAGTGCTTG 3205

Qy       1170 -IleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAl 1189
                  :::    :::    :::    :::|||    |||       |||
Db       3206 GCAGTGCAAGTGCTCATATGAAACACAACTGAACGCACTGCTCAACCAATATGGCGTGCG 3265

Qy       1189 aLysAspThrGlu---AspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrTh 1208
                 :::    |||                |||    |||          |||       |||
Db       3266 CACTGAAGCAGAGCTTGTGACAGAGCATATATGG------TCGCTTCCCAAGTAC----- 3314

Qy       1208 rPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLy 1228
                                          ||||||          :::          |||||
Db       3315 ---------------------AGCAGCAGAAGGCAGGGGGACATAAAGGAGAGGTTGAA 3352

Qy       1228 sAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLysGluMetAr 1248
              ||||          :::    ||||||:::    |||    :::             :::
Db       3353 GAATGCATACTATGCTCTTCACAAGGAGTTT---AGGAGCATTTTCGAAAGCATTGTGAC 3409

Qy       1248 gAspSerLysAspProTyrProValArgValAsnGlnValTyrGlu-----------Ly 1264
                |||    :::                   ::::::    ||||||
Db       3410 AGATCAAACAGAGATCTCTGATGATGAGAAAAGTCGGTTTTACGAGATGAAGGCCTCCGC 3469

Qy       1264 sTrpCysAlaIleThr------ProGluAlaMetAspLysSerGlyAlaAsnTyrAspSe 1282
                |||       :::|||    ||||||    :::    ||||||          :::    ||
Db       3470 TTGGTACCAGGTAACCTACCATCCTGAATGGGTCCAGAAGTCAAGGGAAATGTTCAAGTC 3529

Qy       1282 rLysValIleArgLeu---LeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTr 1301
              |          :::          |||||||||       ||||||    ::::::|||||
Db       3530 TGACTGTGAGGACATGCCAGCAAGGCTTAGCTTTGC-ATGGATCGCGG-TTGAGCACCTG 3587

Qy       1301 pAlaLeuLeuArg 1305
              |       ||||||
Db       3588 GCACGGATTAAGA 3600
```

RESULT 5
AAC63741

```
ID   AAC63741 standard; cDNA; 2816 BP.
XX
AC   AAC63741;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Rice RNA-directed RNA polymerase FIS sequence, SEQ ID NO: 9.
XX
KW   Rice; plant; RNA-directed RNA polymerase; FIS; gene mapping; gene marker;
KW   plant virus resistance; plant breeding; ss.
XX
OS   Oryza sativa.
XX
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
```

```
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT,  Orozco EM,  Weng Z,  Sakai H,  Cahoon RE,  Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   P-PSDB; AAB28532.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 2; Page 45-46; 62pp; English.
XX
CC   The present sequence is one of a number of cDNA molecules which encode
CC   plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC   probes for genetically and physically mapping genes, and as markers for
CC   traits linked to those genes. They are useful for controlling gene
CC   expression and provide mechanisms to engineer plant virus resistance.
CC   They are also useful for plant breeding to develop lines with desired
CC   phenotypes
XX
SQ   Sequence 2816 BP; 782 A; 572 C; 725 G; 736 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:              1.1e-09        Length:          2816
Score:                  317.00         Matches:         208
Percent Similarity:     37.54%         Conservative:    164
Best Local Similarity:  20.99%         Mismatches:      341
Query Match:            4.29%          Indels:          278
DB:                     3              Gaps:            49

US-09-913-878A-2 (1-1402) x AAC63741 (1-2816)

Qy     463 PheArgGlyLysProPhePro-----------------------GluLysProProAsn 474
            |||  |||     ||||||                          :::|||    |||
Db      31 TTCTGGGGGGACAAGTTCCCAGTGTTTGATGCGTGCGGGAGGCTGAAGAAGGCACTGAAT 90

Qy     475 AspVal---------------PheValThrAlaMetThrGlyAsnPheGluSerLysGly 489
              |||                  :::   :::     |||:::    ::::::
Db      91 CGGGTGGCCAGGAACCCCAAACTTCTCTGCAGCAAGGTCGGGGATGACCACGCGGAGGTG 150

Qy     490 SerAlaValValLeuSerAlaValLeuAspTyrAsnProAspAsnSerProThrAlaPro 509
                :::||||||:::                             :::||||||
Db     151 CGGAGCGTGGTGATC-----------------------------ACGCCCACCNAGGCT 180

Qy     510 LeuTyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrArgArg 529
                 |||                |||       |||:::     |||:::   |||
Db     181 TATTGTCTG------------CCTCCAGAAGTGGAGCGCTCAAACCGTGTTCTTCGGCAT 228

Qy     530 PheGly------ProAspArgPhePheGluIleLeuIleProSerPro------------ 543
           :::               ||||||||     :::
Db     229 TACCATGAGGTGGCTGACAGGTTTTTGAGGGTCACTTTATGGACGAGGGTATGCAGGTG 288

Qy     544 ---------------------ThrSerThrSerProSerValProProValValSer--- 555
                                :::   |||:::|||   |||       ::::::|||
Db     289 CTGAACAACAATGTGCTCAACTCCTTCACTGCACCAATTGTCAAAGACTTGATGTCGAAT 348

Qy     556 ---------LysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGln 572
                    ::::::         :::  |||        ||||||    |||
Db     349 TTTTTCCAGCAGAAGACAACGGTGTACAAGCGTGTCAGGATGTTGTTGACGGAGGGTTTC 408

Qy     573 HisSerLeuValGlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLys 592
           |||   :::   |||||||::::::     :::|||    |||        :::
Db     409 CAC---ATGTGTGGAAGGAAGTAC---TCATTTCTCGCATTCTCATCGAAC--------- 453

Qy     593 ProLeuArgGluPheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgVal 612
                          ||||||||                           :::
Db     454 ---------------CAGTTAAGG-----------------------GACAAGTCAGCT 474

Qy     613 HisPhePheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerVal 632
             ||||||||||||
Db     475 TGGTTCTTTGCCGAG--------------------------------------------- 489

Qy     633 ValProAlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAsp 652
                                   :::::::::|||       |||      :::
Db     490 ---------------------GACAGAAAGACA------ACGGTGGAAGCAATTAGGAAG 522

Qy     653 TrpLeuLeuGlnLeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPhe--------- 669
            |||          :::|||   :::                 :::::::::|||
Db     523 TGG---------ATGGACGGTTCA------------CAAGTAAGAATGTTCGAAGATGCT 561

Qy     670 SerArgIleGlnLeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGln 689
            :::|||:::                 |||  |||||||||    :::|||:::   |||   :::
Db     562 GCTCGAATGGGGCAGTGCTTCTCATCCACATATGCAACTGTGACAATGCGGCCGGATGAG 621

Qy     690 IleArgHisHisLysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAsp 709
            :::                  |||::::::                  :::    :::|||
Db     622 GTTGATGAGAGTTTTGATGATGTTGTGCAT---------AATGAGTACATTTTCTCCGAT 672

Qy     710 GlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGly 729
            |||:::|||:::::::::       :::|||       :::    :::    |||    |||
Db     673 GGAATTGGCAAAATTACCCCAGATCTTGCATTGGAAGTTGCCGAGAGGCTGCAGCTGACA 732

Qy     730 Asp---ValProSerAlaValGlnGlyArgPheGlySerAlaLysGly---------Met 745
            |||      |||||||||   |||   ||||||         |||||            :::
Db     733 GATAACCCGCCATCTGCTTATCAGATCAGGTTTGCTGGCTTCAAGGGTGTCATAGCTGTC 792

Qy     746 TrpValIleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln 765
            |||                ||||||          :::                ||||||
Db     793 TGG------------CAAGGACATGGTGATGGGACACGGCTTTTCCTGAGGCCAAGCATG 840

Qy     766 ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerVal 785
            ||||||:::|||    :::   :::                        |||      |||
Db     841 AGGAAGTTTGAGTCTAACCATTTG-----------------------GTGTTAGGGGTG 876

Qy     786 AlaSer------GluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGlu 803
               |||       :::   :::         ||||||   |||::::::      :::|||
Db     877 GTCTCCTGGACAAAGTTCCAGCCAGGATTCTTAAATCGACAGATTATAATATTGCTATCC 936

Qy     804 AspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAsp 823
                                                   :::   |||   :::
Db     937 TCA--------------------------CTGAATGTCCCAGATTCTATCTTTTGGCAA 969

Qy     824 LeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGln 843
            :::|||                               ||||||     :::      |||
Db     970 ATGCAAGAGACCATGCTTTCTAAC---------CTCAACAATATTCTATCAGACAGAGAT 1020

Qy     844 TrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPhe 863
               :::|||   ::::::
Db    1021 GTTGCTTTTGAGGTTTTAACAACTTCATGTGCT--------------------------- 1053

Qy     864 LeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPhe 883
                                         |||       ::::::::::::||||||
Db    1054 ------------------GATGATGGAAACACTGCAGCATTGATGCTCAGTGCTGGCTTT 1095

Qy     884 AspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCys 903
            :::|||:::   :::   :::|||:::   :::         ::::::      :::
Db    1096 GAACCTAGAACTGAACCACACTTGAAAGCAATGCTCTTGGCTATAAGGTCTGCACAATTG 1155

Qy     904 AspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIle 923
                 |||      |||      |||   |||   :::      ::::::
Db    1156 CAGGATCTTTTGGAAAAAGCAAGGATATTTGTGCCAAAGGGAAGGTGGTTGATGGGCTGT 1215

Qy     924 AlaAspPheTrpGlyValLeuGluGlu--------------------------------- 932
              |||       ||||||||||||:::
Db    1216 CTTGATGAGCTTGGGGTTCTTGAGCAAGGGCAGTGCTTTATTCGGGCAACAGTTCCATCA 1275

Qy     933 ---AsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheThr 951
            |||      |||         |||:::|||               :::::::::
Db    1276 TTGAATAGTTATTTTGTTAAGCATGGGTCAAGATTTTCATCAACAGATAAAAACACAGAG 1335
```

```
Qy        952 LeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIle 971
Db       1336 GTCATTTTGGGT---ACTGTGGTAATAGCAAAGAATCCCTGTCTTCATCCAGGGGATGTC 1392
Qy        972 GlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePhe 991
Db       1393 CGCATCCTTGAAGCAGTTGATGTGCCCGAACTGCATCATCTGGTTGATTGTTTGGTGTTC 1452
Qy        992 SerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAsp 1011
Db       1453 CCCCAGAAAGGTGAGAGGCCACATGCTAACGAGGCATCTGGGAGCGATCTTGATGGGGAT 1512
Qy       1012 MetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetPro 1031
Db       1513 CTCTACTTTGTGACATGGGATGAGAAACTTATACCTCCAGGCAAGAAGAGCTGGAACCCT 1572
Qy       1032 LeuGlu---------ProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLys 1048
Db       1573 ATGGACTACTCCCCACCTGAAGCA--------------------------------AAA 1599
Qy       1049 GlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMet 1068
Db       1600 CAACTT--------------------------CCGCGCCAAGTATCTCAACATGATATC 1632
Qy       1069 IleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyr 1088
Db       1633 ATTGATTTCTTCTTAAAGAACATGATAAGTGAGAATCTTGGTAGGATCTGTAATGCTCAT 1692
Qy       1089 LysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSer 1108
Db       1693 GTTGTTCATGCTGATCTTAGCGAGTATGGTGCAATGGATGAGAAGTGTATTCACTTAGCT 1752
Qy       1109 SerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSer 1128
Db       1753 GAGCTAGCAGCAACTGCCGTGGACTTCCCCAAGACTGGC--------------------- 1791
Qy       1129 TrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSer---------LeuProAsp 1145
Db       1792 ------------AAACTTGCGATAATGCCACCACACCTTAAACCAAAAGTCTACCCTGAC 1839
Qy       1146 ProMet--------------------TyrLysSerAspSerTrpLeuGlyArg------ 1156
Db       1840 TTCATGGGAAAGGAAGATGGACAATCTTATAAATCAGAAAAGATTCTTGGAAGGCTTTAT 1899
Qy       1157 ------------------------GlyGluProThrHisIleIleAspTyrLeuLysPhe 1168
Db       1900 CGTTCAATCCAAGAGGCCTCCAATGGT-----------------GATGTGGTGTCACAA 1941
Qy       1169 SerIleAlaArgPro---------------AlaIleAspLysGluLeuGluAlaPheHis 1183
Db       1942 GAGGTTTGCACTCCAAATGATCTGCCTTATGACATAGATCTGGAAGTTCCTGGTGCATCA 2001
Qy       1184 AsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeu 1203
Db       2002 GATTTCCTCGCGAGTGCTTGGCAATGCAAGTGTTCA---------TACGACGCGCAGCTG 2052
Qy       1204 AlaSerTyrTyrThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeu 1223
Db       2053 AGTGCGCTGCTCAGTCAGTAC---------------AGGGTCCGCACTGAAGCTGAACTT 2097
Qy       1224 PheThrThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuValLys 1243
Db       2098 GTGACA------------------GGGCACATAACG------------TTCCTTGTTAAG 2127
Qy       1244 Asn------LysGluMetArgAspSerLysAspProTyrProValArgValAsnGlnVal 1261
Db       2128 AACAGCAGCAAGAAGCAAGGCGACATAAAGGAC------------AGGCTGAAGACTGCT 2175
Qy       1262 TyrGluLysTrpCysAlaIleThrProGluAlaMetAspLysSer---GlyAlaAsnTyr 1280
Db       2176 TAC-----------------------TCTGCACTACGCAAGGAGTTCAAAAGTACCTTT 2211
Qy       1281 AspSerLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThr 1300
Db       2212 GAGAGCATAGCATCGGATCAATGCGAG---------ATTGGTGATGACGAGAAGAATCTG 2262
Qy       1301 TrpAlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPhe 1320
Db       2263 CTGTACGAGATGAAGGCCTCTGCCTGGTACCAGGTGACCTATCAC------CCCAAATGG 2316
Qy       1321 ValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArgPro 1340
Db       2317 GTGGAGAAGTCGAGGGGC---------------------------ATTCTGGGCCCAGAT 2349
Qy       1341 GlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLys 1360
Db       2350 GGTGAGGAAATACCGGCAAGCCTTAGC---------TTTGCATGGATCCCGGTGGAT--- 2397
Qy       1361 LysPheThrLysGlnTyrValAlaArgLeuGlu 1371
Db       2398 ---------------TACTTGGCGCGGATAAAG 2415
```

RESULT 6
AAS17845
ID   AAS17845 standard; cDNA; 3731 BP.
XX
AC   AAS17845;
XX
DT   08-MAY-2002 (first entry)
XX
DE   Tomato RNA-directed RNA polymerase (RdRP) cDNA from clone HF.
XX
KW   Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW   cytostatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW   transgenic plant; transgenic animal; cancer; viral infection;
KW   immunoprecipitation; immunolocalisation; RdRP/HF; mutant; gene therapy;
KW   ss.
XX
OS   Lycopersicon esculentum.
XX
FH   Key   Location/Qualifiers
FT   mutation   replace(31,T)
FT   /*tag= b
FT   mutation   replace(32,G)
FT   /*tag= c
FT   CDS   194..3535
FT   /*tag= a
FT   /product= "RdRP protein"
FT   /EC_number= "2.7.7.48"
FT   mutation   replace(2017,G)
FT   /*tag= d
FT   mutation   replace(2264,T)
FT   /*tag= e
FT   mutation   replace(3597,A)
FT   /*tag= f
FT   mutation   replace(3599,A)
FT   /*tag= g
XX
PN   US2001023067-A1.
XX
PD   20-SEP-2001.
XX
PF   08-FEB-2001; 2001US-00782874.
XX
PR   05-MAR-1997;   97US-00811583.
XX
PA   (WASS/) WASSENEGGER M.
PA   (RIED/) RIEDEL L.
XX
PI   Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
DR   WPI; 2001-595798/67.
XX
PT   New nucleic acid molecule encoding a polypeptide having the enzymatic
PT   activity of RNA-directed RNA polymerase, for modulating gene expression
PT   and treating cancer and virus infection in human and animals.

```
XX
PS   Example 3; Page; 34pp; English.
XX
CC   This sequence represents a cDNA encoding the tomato RNA-directed RNA
CC   polymerase (RdRP/HF) of the invention. The invention comprises the
CC   nucleic acid and protein sequences of RdRP. The protein of the invention
CC   can catalyse in vitro transcription of short single stranded RNAs into
CC   DNA molecules, this transcription can be either primed by RNA or DNA
CC   oligonucleotides or be unprimed. The protein may have cytostatic or
CC   virucide activities. The sequences of the invention may be used in gene
CC   therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA
CC   sequence and a template nucleic acid molecule derived from a nucleic acid
CC   molecule which causes a disease are useful for treating a disease caused
CC   by the undesired expression or overexpression of a nucleic acid molecule
CC   in a human, rat or mouse, by administering the molecules. This system can
CC   be used in the preparation of a pharmaceutical composition and for
CC   inhibiting expression of any desired gene by transferring the RdRP system
CC   to organisms that either lack a comparable mechanism or do not
CC   sufficiently express their own RdRP. An antibody or an antagonist or
CC   inhibitor to the protein are useful for inhibiting RNA directed RNA
CC   synthesis and for ensuring stable heterologous, gene expression in
CC   transgenic organisms. The sequence is useful for probes and/or for the
CC   control of gene expression, as primers for amplification of nucleic acid
CC   molecules and as tools for the detection of expression of the cDNA
CC   molecules. Additionally, nucleotide and protein sequences are useful for
CC   suppression of undesired gene expression in humans and animals. The RdRP
CC   is useful as a therapeutic agent for the control of cancer and virus
CC   infection in humans and animals and the antibody is useful for
CC   immunoprecipitation or immunolocalisation of the protein, identification
CC   of polypeptides interacting with it and screening expression libraries.
CC   Note: This sequence is not shown in the specification but was created by
CC   the indexer from the wild type RdRP sequence shown in AAS17837 and the
CC   information shown in table 1
XX
SQ   Sequence 3731 BP; 1062 A; 672 C; 848 G; 1149 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.3e-08        Length:         3731
Score:                  297.50         Matches:        189
Percent Similarity:     35.78%         Conservative:   137
Best Local Similarity:  20.75%         Mismatches:     339
Query Match:            4.02%          Indels:         247
DB:                     4              Gaps:           35

US-09-913-878A-2 (1-1402) x AAS17845 (1-3731)

Qy      562 GluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArg 581
             |||::::::        :::|||   :::          |||   |||     :::|||
Db     1349 GAAGACATAGATAACTTTCTTCGTGTT---------TCTTTTGTTGATGAGGAGTGGGAG 1399

Qy      582 AlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAla 601
                :::::   |||
Db     1400 AAACTGTATTCTACAGACTTA--------------------------------------- 1420

Qy      602 GluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIleThr 621
                  |||||                                  |||   |||||
Db     1421 ---TTACCAAAA------------------------------GCAAGTACTGGAAGTGGT 1447

Qy      622 PheArgProAspValPheLys------------------ThrArgSerValValProAla 635
               |||     :::::::::::                   :::      |||
Db     1448 GTCAGGACAAACATCTATGAGAGGATCTTATCAACTCTGCGGAAAGGCTTTGTAATTGGT 1507

Qy      636 GluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeu 655
            :::::        |||               |||      ||||||:::
Db     1508 GATAAAAAATTTGAA---TTTCTTGCATTTTCATCGAGCCAGTTG--------------- 1549

Qy      656 GlnLeuAspAsnAsnThrTrp-----------GlnProHisLeuLysLeu--------- 668
                  ||||||:::  |||            :::|||   |||
Db     1550 ---CGGGATAATTCAGTGTGGATGTTTGCATCAAGACCTGGCCTTACTGCAAATGATATA 1606

Qy      669 -----------------PheSerArgIle-----------------------GlnLeu 674
                             ||||||:::|||                        :::|||
Db     1607 AGAGCTTGGATGGGTGATTTTTCGCAGATCAAGAATGTCGCAAAATATGCTGCCAGACTT 1666

Qy      675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLys 694
            |||   |||             :::      ||||||              :::        |||:::
Db     1667 GGTCAATCTTTTGGTTCCTCCAGAGAGACTTTG---------AGTGTTCTTAGGCATGAG 1717

Qy      695 ThrAspLeuLeu---------SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
               :::::::::                  ||||||      |||    :::||||||:::
Db     1718 ATTGAAGTTATTCCCGATGTAAAGGTTCATGGAACCAGCTATGTCTTTTCTGATGGAATT 1777

Qy      712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal 731
            |||::::::|||            |||   |||:::               ||||||
Db     1778 GGTAAAATATCTGGTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCTTCAATATACC 1837

Qy      732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAsp 751
            |||||||||   |||   |||:::|||       ||||||:::         :::|||   |||
Db     1838 CCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGGTGTTGATCCGGAT 1897

Qy      752 AspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAsp 771
               :::            :::         |||      |||:::|||   |||
Db     1898 TCATCAATGAAG---------TTGTCTTTGAGAAAGAGCATGTCGAAATATGAATCAGAC 1948

Qy      772 PheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791
               :::                  |||:::|||         |||:::   :::
Db     1949 AACATA---------------AAGTTAGATGTCCTTGGA---TGGAGCAAATATCAGCCT 1990

Qy      792 AlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLys 811
               ||||||   ||||||:::   :::|||
Db     1991 TGTTATCTTAATCGTCAACTGATTACTCTCTTGTCT------------------------ 2026

Qy      812 MetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGln 831
                 :::   |||   ::::::            |||::::::      :::|||
Db     2027 ---ACACTTGGAGTGAAAGATGAAGTTCTCGAACAGAAGCAAAAGGAAGCTGTAGATCAG 2083

Qy      832 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 851
               |||:::                        |||      |||      |||
Db     2084 CTTGATGCTATCTTGCATGATTCTTTGAAGGCACAGGAGGCTTTGGAATTGATGTCT--- 2140

Qy      852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
                                                            |||
Db     2141 ------------------------------------------------CCTGGAGAGAAC 2152

Qy      872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeu 891
                  |||      ::::::|||   |||:::   |||      :::   :::|||
Db     2153 ACTAATATTCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCCTTTCTT 2212

Qy      892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
               :::         :::      |||         |||:::::::::
Db     2213 TCAATGATGTTGCAAACCTTCCGCGCATCCAAGTTGCTCGACTTGCGGACTAGATCAAGA 2272

Qy      912 IleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGlu 931
            |||   :::               :::         |||            ||||||
Db     2273 ATATTTATTCCAAATGGAAGAACAATGATGGGATGTTTGGATGAATCCAGAACCTTGGAA 2332

Qy      932 GluAsnGluValHisValGlyPheSer------------SerLysPheArgAspGluGlu 947
               :::|||   |||   |||:::                     :::|||   |||:::
Db     2333 TATGGTCAGGTGTTTGTTCAGTTTACTGGTGCTGGACATGGAGAGTTTTCTGACGATTTA 2392

Qy      948 GluSerPhe------------------------ThrLeuLeuSerAspCysAspValLeu 959
               |||                                :::         :::|||:::
Db     2393 CATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAATGTGGTT 2452

Qy      960 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 979
            ||||||::::::|||            |||   ||||||:::   ::::::||||||
Db     2453 GTTGCAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTTAAAGGCTGTAAATGTT 2512

Qy      980 ProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeu 999
               ||||||   :::   |||   ::::::|||      ||||||      |||
Db     2513 CGAGCGCTGCACCACATGGTAGATTGTGTTGTATTCCCTCAGAAAGGAAAAAGACCTCAT 2572

Qy     1000 AlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspPro 1019
               :::   ||||||   |||   |||||||||:::   :::||||||||||||
Db     2573 CCGAATGAATGTTCTGGGAGTGATTTGGATGGGGATATCTACTTTGTTTGCTGGGATCAA 2632
```

```
Qy       1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
              :::::::::                                 |||        |||     |||
Db       2633 GACATGATCCCGCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCCAGCATACAG 2692

Qy       1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
              |||       |||
Db       2693 TTGGACCATGAT------------------------------------------------ 2704

Qy       1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 1079
                          |||  :::  :::::::::    |||          :::
Db       2705 ------------GTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATTGTGAAT 2752

Qy       1080 AsnPheLeuGlyMetCysThrAsn------------TyrLysGluArgLeuCysTyrIle 1095
              :::    ||||||:::       |||             :::   :::|||
Db       2753 GACAGTTTGGGAATCATAGCAAATGCCCATGTCGTATTTGCAGACAGA---------GAA 2803

Qy       1096 AsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVal 1115
                 :::         :::    |||          |||:::    |||          |||
Db       2804 CCTGATATGGCCATGAGTGATCCATGCAAAAAACTTGCTGAGCTCTTTTCAATTGCAGTG 2863

Qy       1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 1135
              |||       |||   |||:::                |||      ||||||||
Db       2864 GACTTTCCAAAGACTGGTGTTCCCGCTGAAATACCATCT------CAGTTGCGC------ 2911

Qy       1136 LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTrpLeuGl 1155
                                                |||          ||| ||||||:::
Db       2912 -----------------------------CCTAAAGAATACCCAGACTTCATGGATAAG 2941

Qy       1155 yArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIl 1175
               |||                                          ||||||||||
Db       2942 CCGG--------------------------------------ACAAGACCAGCTAT 2959

Qy       1175 eAspLysGlu------LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGl 1193
              |  ::::::         ||||||:::|||              :::|||
Db       2960 ATCTCAGAAAGAGTTATTGGAAAGCTTTTC-AGGAAAGTGAAG----------------- 3001

Qy       1193 uAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGluIl 1213
               |||  |||          |||        :::|||       |||||   ::::::::
Db       3002 -GACAAAGCA------------CCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGT 3048

Qy       1213 eSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGl 1233
              ::::   :::|||        :::   :::              ::::::     |||    ||
Db       3049 TGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACATTGACGA 3108

Qy       1233 uVal---------GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSe 1250
              |                 :::     ||||||              :::     ::::::   :::
Db       3109 AGCTTTTGACTACAAAACTGAATAT---------------GACAACAAGCTGGGTAATTT 3153

Qy       1250 rLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrPr 1270
                   |||    |||   ::::::                                    |||
Db       3154 AATGGACTACTATGGCATAAAAACAGAG------------------GCTGAAATACTTAG 3195

Qy       1270 oGluAlaMetAspLysSerGlyAlaAsnTyrAsp-----------------------Se 1282
                  :::     |||:::           :::|||                           ||
Db       3196 TGGTGGCATTATGAAGGCATCAAAAACTTTTGACCGCAGAAAAGATGCTGAGGCCATTAG 3255

Qy       1282 rLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp-- 1301
              |     :::|||   |||                       :::|||          |||
Db       3256 TGTTGCTGTGAGGGCCTTGAGG------------------AAGGAGGCAAGAGCCTGGTT 3297

Qy       1302 ------------------------AlaLeuLeuArgAlaSerThrAlaPheLysLeuTy 1313
                                        |||   :::||||||      :::    :::
Db       3298 CAAGAGGCGTAATGATATAGATGACATGTTACCAAAGGCTTCGGCTTGGTACCACGTTAC 3357

Qy       1313 rTyrHisLysSer---------------------------------------ProLy 1319
                ||||||   :::
Db       3358 ATATCATCCTACATATTGGGGTTGCTACAATCAGGGGTTGAAAAGAGCTCATTTCATTAG 3417

Qy       1319 sPheValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerAr 1339
                 |||   |||   :::       :::   :::          |||   :::      :::||
Db       3418 CTTTCCCTGGTGTGTTTATGACCAGCTAATCCAGATTAAGAAGGACAAAGCACGTAACAG 3477

Qy       1339 gProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAs 1359
                 ||||                  :::|||            :::         :::
Db       3478 GCCAGTTCTCAACTTGTCATCTCTCAGGGCTCAACTGAGTCACAGATTAGTG-------- 3529

Qy       1359 pLysLysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAs 1379
                                                 |||:::    |||      |||
Db       3530 -----------------------------TTGAAATGAGATTCCAGTCGAGCGTTAAG 3558

Qy       1379 pProGluValTyrGluValLeuGlyAspAsp 1389
                   :::|||    |||:::|||    |||
Db       3559 CTGATATATATATAATGTAATAGGGTGTGAT 3589
```

RESULT 7
AAS17837

```
ID   AAS17837 standard; cDNA; 3731 BP.
XX
AC   AAS17837;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Tomato RNA-directed RNA polymerase (RdRP) cDNA.
XX
KW   Tomato; RdRP; RNA-directed RNA polymerase; in vitro transcription;
KW   cytostatic; virucide; RNA synthesis inhibitor; antibody; immunogen;
KW   transgenic plant; transgenic animal; cancer; viral infection;
KW   immunoprecipitation; immunolocalisation; ss; gene therapy.
XX
OS   Lycopersicon esculentum.
XX
FH   Key             Location/Qualifiers
FT   CDS             194. .3538
FT                   /*tag=  a
FT                   /product= "RdRP protein"
FT                   /EC_number= "2.7.7.48"
XX
PN   US2001023067-A1.
XX
PD   20-SEP-2001.
XX
PF   08-FEB-2001; 2001US-00782874.
XX
PR   05-MAR-1997;   97US-00811583.
XX
PA   (WASS/) WASSENEGGER M.
PA   (RIED/) RIEDEL L.
XX
PI   Wassenegger M,  Riedel L,  Schiebel W,  Sanger HL;
XX
DR   WPI; 2001-595798/67.
DR   P-PSDB; AAU10006.
XX
PT   New nucleic acid molecule encoding a polypeptide having the enzymatic
PT   activity of RNA-directed RNA polymerase, for modulating gene expression
PT   and treating cancer and virus infection in human and animals.
XX
PS   Claim 1; Page 15-19; 34pp; English.
XX
CC   This sequence represents a cDNA encoding the tomato RNA-directed RNA
CC   polymerase (RdRP) protein of the invention. The invention comprises the
CC   nucleic acid and protein sequences of RdRP. The protein of the invention
CC   can catalyse in vitro transcription of short single stranded RNAs into
CC   DNA molecules, this transcription can be either primed by RNA or DNA
CC   oligonucleotides or be unprimed. The protein may have cytostatic or
CC   virucide activities. The sequences of the invention may be used in gene
CC   therapy or as an RNA directed RNA synthesis inhibitor. The RdRP cDNA
CC   sequence and a template nucleic acid molecule derived from a nucleic acid
CC   molecule which causes a disease are useful for treating a disease caused
CC   by the undesired expression or overexpression of a nucleic acid molecule
CC   in a human, rat or mouse, by administering the molecules. This system can
CC   be used in the preparation of a pharmaceutical composition and for
CC   inhibiting expression of any desired gene by transferring the RdRP system
CC   to organisms that either lack a comparable mechanism or do not
CC   sufficiently express their own RdRP. An antibody or an antagonist or
CC   inhibitor to the protein are useful for inhibiting RNA directed RNA
```

```
CC   synthesis and for ensuring stable heterologous, gene expression in
CC   transgenic organisms. The sequence is useful for probes and/or for the
CC   control of gene expression, as primers for amplification of nucleic acid
CC   molecules and as tools for the detection of expression of the cDNA
CC   molecules. Additionally, nucleotide and protein sequences are useful for
CC   suppression of undesired gene expression in humans and animals. The RdRP
CC   is useful as a therapeutic agent for the control of cancer and virus
CC   infection in humans and animals and the antibody is useful for
CC   immunoprecipitation or immunolocalisation of the protein, identification
CC   of polypeptides interacting with it and screening expression libraries
XX
SQ   Sequence 3731 BP; 1064 A; 669 C; 849 G; 1149 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               2.3e-08     Length:         3731
Score:                   297.50      Matches:        189
Percent Similarity:      35.78%      Conservative:   137
Best Local Similarity:   20.75%      Mismatches:     339
Query Match:             4.02%       Indels:         247
DB:                      4           Gaps:           35

US-09-913-878A-2 (1-1402) x AAS17837 (1-3731)

Qy    562 GluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArg 581
Db   1349 GAAGACATAGATAACTTTCTTCGTGTT---------TCTTTTGTTGATGAGGAGTGGGAG 1399

Qy    582 AlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAla 601
Db   1400 AAACTGTATTCTACAGACTTA--------------------------------------- 1420

Qy    602 GluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIleThr 621
Db   1421 ---TTACCAAAA------------------------------GCAAGTACTGGAAGTGGT 1447

Qy    622 PheArgProAspValPheLys------------------ThrArgSerValValProAla 635
Db   1448 GTCAGGACAAACATCTATGAGAGGATCTTATCAACTCTGCGGAAAGGCTTTGTAATTGGT 1507

Qy    636 GluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeu 655
Db   1508 GATAAAAAATTTGAA---TTTCTTGCATTTTCATCGAGCCAGTTG--------------- 1549

Qy    656 GlnLeuAspAsnAsnThrTrp------------GlnProHisLeuLysLeu--------- 668
Db   1550 ---CGGGATAATTCAGTGTGGATGTTTGCATCAAGACCTGGCCTTACTGCAAATGATATA 1606

Qy    669 ------------------PheSerArgIle------------------------GlnLeu 674
Db   1607 AGAGCTTGGATGGGTGATTTTTCGCAGATCAAGAATGTCGCAAAATATGCTGCCAGACTT 1666

Qy    675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLys 694
Db   1667 GGTCAATCTTTTGGTTCCTCCAGAGAGACTTTG---------AGTGTTCTTAGGCATGAG 1717

Qy    695 ThrAspLeuLeu---------SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
Db   1718 ATTGAAGTTATTCCCGATGTAAAGGTTCATGGAACCAGCTATGTCTTTTCTGATGGAATT 1777

Qy    712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal 731
Db   1778 GGTAAAATATCTGGTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCTTCAATATACC 1837

Qy    732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAsp 751
Db   1838 CCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGGTGTTGATCCGGAT 1897

Qy    752 AspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAsp 771
Db   1898 TCATCAATGAAG---------TTGTCTTTGAGAAAGAGCATGTCGAAATATGAATCAGAC 1948

Qy    772 PheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791
Db   1949 AACATA---------------AAGTTAGATGTCCTTGGA---TGGAGCAAATATCAGCCT 1990

Qy    792 AlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLys 811
Db   1991 TGTTATCTTAATCGTCAACTGATTACGCTCTTGTCT------------------------ 2026

Qy    812 MetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGln 831
Db   2027 ---ACACTTGGAGTGAAAGATGAAGTTCTCGAACAGAAGCAAAAGGAAGCTGTAGATCAG 2083

Qy    832 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 851
Db   2084 CTTGATGCTATCTTGCATGATTCTTTGAAGGCACAGGAGGCTTTGGAATTGATGTCT--- 2140

Qy    852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
Db   2141 ------------------------------------------------CCTGGAGAGAAC 2152

Qy    872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeu 891
Db   2153 ACTAATATTCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCCTTTCTT 2212

Qy    892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
Db   2213 TCAATGATGTTGCAAACCTTCCGCGCATCCAAGTTGCTCGATTTGCGGACTAGATCAAGA 2272

Qy    912 IleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGlu 931
Db   2273 ATATTTATTCCAAATGGAAGAACAATGATGGGATGTTTGGATGAATCCAGAACCTTGGAA 2332

Qy    932 GluAsnGluValHisValGlyPheSer------------SerLysPheArgAspGluGlu 947
Db   2333 TATGGTCAGGTGTTTGTTCAGTTTACTGGTGCTGGACATGGAGAGTTTTCTGACGATTTA 2392

Qy    948 GluSerPhe------------------------ThrLeuLeuSerAspCysAspValLeu 959
Db   2393 CATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAATGTGGTT 2452

Qy    960 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 979
Db   2453 GTTGCAAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTTAAAGGCTGTAAATGTT 2512

Qy    980 ProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeu 999
Db   2513 CGAGCGCTGCACCACATGGTAGATTGTGTTGTATTCCCTCAGAAAGGAAAAAGACCTCAT 2572

Qy   1000 AlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspPro 1019
Db   2573 CCGAATGAATGTTCTGGGAGTGATTTGGATGGGGATATCTACTTTGTTTGCTGGGATCAA 2632

Qy   1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
Db   2633 GACATGATCCCGCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCCAGCATACAG 2692

Qy   1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
Db   2693 TTGGACCATGAT------------------------------------------------ 2704

Qy   1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 1079
Db   2705 ------------GTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATTGTGAAT 2752

Qy   1080 AsnPheLeuGlyMetCysThrAsn------------TyrLysGluArgLeuCysTyrIle 1095
Db   2753 GACAGTTTGGGAATCATAGCAAATGCCCATGTCGTATTTGCAGACAGA---------GAA 2803

Qy   1096 AsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVal 1115
Db   2804 CCTGATATGGCCATGAGTGATCCATGCAAAAAACTTGCTGAGCTCTTTTCAATTGCAGTG 2863

Qy   1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 1135
Db   2864 GACTTTCCAAAGACTGGTGTTCCCGCTGAAATACCATCT------CAGTTGCGC------ 2911
```

```
Qy       1136 LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTrpLeuGl 1155
                                                |||         ||| ||||||:::
Db       2912 -----------------------------CCTAAAGAATACCCAGACTTCATGGATAAG 2941

Qy       1155 yArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIl 1175
               |||                                          ||||||||||
Db       2942 CCGG-----------------------------------------ACAAGACCAGCTAT 2959

Qy       1175 eAspLysGlu------LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGl 1193
              |  ::::::       ||||||:::|||         :::|||
Db       2960 ATCTCAGAAAGAGTTATTGGAAAGCTTTTC-AGGAAAGTGAAG---------------- 3001

Qy       1193 uAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGluIl 1213
                |||  |||         |||      :::|||     ||||||  ::::::::
Db       3002 -GACAAAGCA-----------CCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGT 3048

Qy       1213 eSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGl 1233
              ::::    :::|||  :::   :::              ::::::   |||   ||
Db       3049 TGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACATTGACGA 3108

Qy       1233 uVal---------GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSe 1250
              |              :::   ||||||               :::  ::::::  :::
Db       3109 AGCTTTTGACTACAAAACTGAATAT--------------GACAACAAGCTGGGTAATTT 3153

Qy       1250 rLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrPr 1270
                 |||   |||   ::::::                              |||
Db       3154 AATGGACTACTATGGCATAAAAACAGAG------------------GCTGAAATACTTAG 3195

Qy       1270 oGluAlaMetAspLysSerGlyAlaAsnTyrAsp------------------------Se 1282
                 :::   |||:::      :::|||                             ||
Db       3196 TGGTGGCATTATGAAGGCATCAAAAACTTTTGACCGCAGAAAAGATGCTGAGGCCATTAG 3255

Qy       1282 rLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp-- 1301
              |       :::|||   |||                    :::|||          |||
Db       3256 TGTTGCTGTGAGGGCCTTGAGG------------------AAGGAGGCAAGAGCCTGGTT 3297

Qy       1302 ------------------------AlaLeuLeuArgAlaSerThrAlaPheLysLeuTy 1313
                                        |||   :::||||||      :::   :::
Db       3298 CAAGAGGCGTAATGATATAGATGACATGTTACCAAAGGCTTCGGCTTGGTACCACGTTAC 3357

Qy       1313 rTyrHisLysSer----------------------------------------ProLy 1319
               ||||||   :::
Db       3358 ATATCATCCTACATATTGGGGTTGCTACAATCAGGGGTTGAAAAGAGCTCATTTCATTAG 3417

Qy       1319 sPheValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerAr 1339
                 |||   |||   :::            :::   |||   :::      :::||
Db       3418 CTTTCCCTGGTGTGTTTATGACCAGCTAATCCAGATTAAGAAGGACAAAGCACGTAACAG 3477

Qy       1339 gProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAs 1359
              ||||           \  :::|||               :::       :::
Db       3478 GCCAGTTCTCAACTTGTCATCTCTCAGGGCTCAACTGAGTCACAGATTAGTG-------- 3529

Qy       1359 pLysLysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAs 1379
                                              |||:::   |||    |||
Db       3530 ------------------------------TTGAAATGAGATTCCAGTCGAGCGTTAAG 3558

Qy       1379 pProGluValTyrGluValLeuGlyAspAsp 1389
                    :::|||   |||:::|||   |||
Db       3559 CTGATATATATATAATGTAATAGGGTGTGAT 3589
```

RESULT 8
AAD04370
ID   AAD04370 standard; cDNA; 3731 BP.
XX
AC   AAD04370;
XX
DT   04-JUL-2001 (first entry)
XX
DE   Tomato RNA-directed RNA polymerase (RdRP) cDNA.
XX
KW   Tomato; gene therapy; RNA-directed RNA polymerase; RdRP; gene expression;
KW   transgenic plant; tissue culture; plant breeding; therapy; C-protein; ss.
XX
OS   Lycopersicon esculentum.
XX
FH   Key   Location/Qualifiers
FT   CDS   194..3538
FT   /*tag= a
FT   /product= "Tomato C-protein having RNA-directed RNA
FT   polymerase (RdRP) activity"
XX
PN   US6218142-B1.
XX
PD   17-APR-2001.
XX
PF   05-MAR-1997;   97US-00811583.
XX
PR   05-MAR-1997;   97US-00811583.
XX
PA   (WASS/) WASSENEGGER M.
PA   (RIED/) RIEDEL L.
XX
PI   Wassenegger M, Riedel L, Schiebel W, Sanger HL;
XX
DR   WPI; 2001-289830/30.
DR   P-PSDB; AAE00897.
XX
PT   New nucleic acid molecules encoding polypeptides with RNA-directed RNA
PT   polymerase enzymatic activity, useful in modulating gene expression in
PT   plants, humans and animals, as well as in plant cell/tissue cultures or
PT   plant breeding.
XX
PS   Claim 1; Col 25-34; 31pp; English.
XX
CC   The present sequence is a cDNA encoding tomato C-protein having RNA-
CC   directed RNA polymerase (RdRP) activity. This protein is capable of RNA-
CC   directed RNA synthesis, thus using RNA as a template for synthesising
CC   complementary RNA molecules. RdRP nucleic acid is useful for modulating
CC   gene expression in plants, humans and animals. This may lead to various
CC   physiological, developmental and/or morphological changes. Transgenic
CC   plants containing RdRP nucleic acid is especially useful in plant cell or
CC   tissue cultures and in plant breeding. RdRP is useful in gene therapy,
CC   particularly for treating a disease that is caused by the undesirable
CC   expression or overexpression of a gene
XX
SQ   Sequence 3731 BP; 1064 A; 669 C; 849 G; 1149 T; 0 U; 0 Other;

```
Alignment Scores:
Pred. No.:               2.3e-08       Length:          3731
Score:                   297.50        Matches:         189
Percent Similarity:      35.78%        Conservative:    137
Best Local Similarity:   20.75%        Mismatches:      339
Query Match:             4.02%         Indels:          247
DB:                      4             Gaps:            35
```

US-09-913-878A-2 (1-1402) x AAD04370 (1-3731)

```
Qy        562 GluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArg 581
              |||::::::      :::|||   :::             |||   |||      :::|||
Db       1349 GAAGACATAGATAACTTTCTTCGTGTT---------TCTTTTGTTGATGAGGAGTGGGAG 1399

Qy        582 AlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAla 601
                    ::::::  |||
Db       1400 AAACTGTATTCTACAGACTTA--------------------------------------- 1420

Qy        602 GluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIleThr 621
                    ||||||                                   |||   ||||||
Db       1421 ---TTACCAAAA------------------------------GCAAGTACTGGAAGTGGT 1447

Qy        622 PheArgProAspValPheLys------------------ThrArgSerValValProAla 635
                 |||   :::::::::::                      :::      |||
Db       1448 GTCAGGACAAACATCTATGAGAGGATCTTATCAACTCTGCGGAAAGGCTTTGTAATTGGT 1507

Qy        636 GluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeu 655
              ::::::      |||               |||      ||||||:::
```

```
Db     1508 GATAAAAAATTTGAA---TTTCTTGCATTTTCATCGAGCCAGTTG-------------- 1549
Qy      656 GlnLeuAspAsnAsnThrTrp-----------GlnProHisLeuLysLeu-------- 668
Db     1550 ---CGGGATAATTCAGTGTGGATGTTTGCATCAAGACCTGGCCTTACTGCAAATGATATA 1606
Qy      669 -----------------PheSerArgIle----------------------GlnLeu 674
Db     1607 AGAGCTTGGATGGGTGATTTTTCGCAGATCAAGAATGTCGCAAAATATGCTGCCAGACTT 1666
Qy      675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLys 694
Db     1667 GGTCAATCTTTTGGTTCCTCCAGAGAGACTTTG---------AGTGTTCTTAGGCATGAG 1717
Qy      695 ThrAspLeuLeu---------SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
Db     1718 ATTGAAGTTATTCCCGATGTAAAGGTTCATGGAACCAGCTATGTCTTTTCTGATGGAATT 1777
Qy      712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal 731
Db     1778 GGTAAAATATCTGGTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCTTCAATATACC 1837
Qy      732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAsp 751
Db     1838 CCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGGTGTTGATCCGGAT 1897
Qy      752 AspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAsp 771
Db     1898 TCATCAATGAAG---------TTGTCTTTGAGAAAGAGCATGTCGAAATATGAATCAGAC 1948
Qy      772 PheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791
Db     1949 AACATA---------------AAGTTAGATGTCCTTGGA---TGGAGCAAATATCAGCCT 1990
Qy      792 AlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLys 811
Db     1991 TGTTATCTTAATCGTCAACTGATTACGCTCTTGTCT------------------------ 2026
Qy      812 MetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGln 831
Db     2027 ---ACACTTGGAGTGAAAGATGAAGTTCTCGAACAGAAGCAAAAGGAAGCTGTAGATCAG 2083
Qy      832 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 851
Db     2084 CTTGATGCTATCTTGCATGATTCTTTGAAGGCACAGGAGGCTTTGGAATTGATGTCT--- 2140
Qy      852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
Db     2141 ------------------------------------------------CCTGGAGAGAAC 2152
Qy      872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeu 891
Db     2153 ACTAATATTCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCCTTTCTT 2212
Qy      892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
Db     2213 TCAATGATGTTGCAAACCTTCCGCGCATCCAAGTTGCTCGATTTGCGGACTAGATCAAGA 2272
Qy      912 IleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGlu 931
Db     2273 ATATTTATTCCAAATGGAAGAACAATGATGGGATGTTTGGATGAATCCAGAACCTTGGAA 2332
Qy      932 GluAsnGluValHisValGlyPheSer-----------SerLysPheArgAspGluGlu 947
Db     2333 TATGGTCAGGTGTTTGTTCAGTTTACTGGTGCTGGACATGGAGAGTTTTCTGACGATTTA 2392
Qy      948 GluSerPhe-----------------------ThrLeuLeuSerAspCysAspValLeu 959
Db     2393 CATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAATGTGGTT 2452
Qy      960 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 979
Db     2453 GTTGCAAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTTAAAGGCTGTAAATGTT 2512
```

```
Qy      980 ProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeu 999
Db     2513 CGAGCGCTGCACCACATGGTAGATTGTGTTGTATTCCCTCAGAAAGGAAAAAGACCTCAT 2572
Qy     1000 AlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspPro 1019
Db     2573 CCGAATGAATGTTCTGGGAGTGATTTGGATGGGGATATCTACTTTGTTTGCTGGGATCAA 2632
Qy     1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
Db     2633 GACATGATCCCGCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCCAGCATACAG 2692
Qy     1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
Db     2693 TTGGACCATGAT------------------------------------------------ 2704
Qy     1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 1079
Db     2705 ------------GTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATTGTGAAT 2752
Qy     1080 AsnPheLeuGlyMetCysThrAsn-----------TyrLysGluArgLeuCysTyrIle 1095
Db     2753 GACAGTTTGGGAATCATAGCAAATGCCCATGTCGTATTTGCAGACAGA---------GAA 2803
Qy     1096 AsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVal 1115
Db     2804 CCTGATATGGCCATGAGTGATCCATGCAAAAAACTTGCTGAGCTCTTTTCAATTGCAGTG 2863
Qy     1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 1135
Db     2864 GACTTTCCAAAGACTGGTGTTCCCGCTGAAATACCATCT------CAGTTGCGC------ 2911
Qy     1136 LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTrpLeuGl 1155
Db     2912 -----------------------------CCTAAAGAATACCCAGACTTCATGGATAAG 2941
Qy     1155 yArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIl 1175
Db     2942 CCGG----------------------------------------ACAAGACCAGCTAT 2959
Qy     1175 eAspLysGlu------LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGl 1193
Db     2960 ATCTCAGAAAGAGTTATTGGAAAGCTTTTC-AGGAAAGTGAAG---------------- 3001
Qy     1193 uAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGluIl 1213
Db     3002 -GACAAAGCA-----------CCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGT 3048
Qy     1213 eSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGl 1233
Db     3049 TGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACATTGACGA 3108
Qy     1233 uVal---------GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSe 1250
Db     3109 AGCTTTTGACTACAAAACTGAATAT---------------GACAACAAGCTGGGTAATTT 3153
Qy     1250 rLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrPr 1270
Db     3154 AATGGACTACTATGGCATAAAAACAGAG-----------------GCTGAAATACTTAG 3195
Qy     1270 oGluAlaMetAspLysSerGlyAlaAsnTyrAsp----------------------Se 1282
Db     3196 TGGTGGCATTATGAAGGCATCAAAAACTTTTGACCGCAGAAAAGATGCTGAGGCCATTAG 3255
Qy     1282 rLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp-- 1301
Db     3256 TGTTGCTGTGAGGGCCTTGAGG-----------------AAGGAGGCAAGAGCCTGGTT 3297
Qy     1302 ------------------------AlaLeuLeuArgAlaSerThrAlaPheLysLeuTy 1313
Db     3298 CAAGAGGCGTAATGATATAGATGACATGTTACCAAAGGCTTCGGCTTGGTACCACGTTAC 3357
```

```
Qy       1313 rTyrHisLysSer----------------------------------------ProLy 1319
                |||||    :::
Db       3358 ATATCATCCTACATATTGGGGTTGCTACAATCAGGGGTTGAAAAGAGCTCATTTCATTAG 3417

Qy       1319 sPheValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerAr 1339
               |||   |||   :::       :::  :::         |||  :::     :::||
Db       3418 CTTTCCCTGGTGTGTTTATGACCAGCTAATCCAGATTAAGAAGGACAAAGCACGTAACAG 3477

Qy       1339 gProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAs 1359
               ||||                  :::|||        :::        :::
Db       3478 GCCAGTTCTCAACTTGTCATCTCTCAGGGCTCAACTGAGTCACAGATTAGTG-------- 3529

Qy       1359 pLysLysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAs 1379
                                             |||:::   |||    |||
Db       3530 ------------------------------TTGAAATGAGATTCCAGTCGAGCGTTAAG 3558

Qy       1379 pProGluValTyrGluValLeuGlyAspAsp 1389
                    :::|||   |||:::|||   |||
Db       3559 CTGATATATATATAATGTAATAGGGTGTGAT 3589

RESULT 9
AAH77703
ID   AAH77703 standard; DNA; 6863 BP.
XX
AC   AAH77703;
XX
DT   13-NOV-2001  (first entry)
XX
DE   Nucleotide sequence of the SGS2 gene.
XX
KW   SGS2; RNA-dependent RNA polymerase; transgene silencing;
KW   transgene stability; crop plant; viral resistance; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   promoter        1. .850
FT                   /*tag= a
FT   exon            851. .3564
FT                   /*tag= b
FT                   /number= 1
FT   intron          3565. .3986
FT                   /*tag= c
FT                   /number= 1
FT   exon            3987. .4862
FT                   /*tag= d
FT                   /number= 2
FT   terminator      4863. .6863
FT                   /*tag= e
XX
PN   FR2804128-A1.
XX
PD   27-JUL-2001.
XX
PF   26-JAN-2000; 2000FR-00001007.
XX
PR   26-JAN-2000; 2000FR-00001007.
XX
PA   (RHOB-) RHOBIO.
PA   (INRG ) INST NAT RECH AGRONOMIQUE.
XX
PI   Beclin C,  Elmayan T,  Mourrain P,  Vaucheret H;
XX
DR   WPI; 2001-543303/61.
XX
PT   A new plant SGS2 gene involved in encoding an RNA-dependent RNA
PT   polymerase and in transgene silencing, increases transgene stability and
PT   expression in transgenic plants when it is inactivated.
XX
PS   Claim 2; Page 30-32; 46pp; French.
XX
CC   The present sequence encodes a plant SGS2 polypeptide. SGS2 is an RNA-
CC   dependent RNA polymerase and is involved in transgene silencing.
CC   Inactivation of SGS2 is used to increase transgene stability and
CC   expression in plants, particularly crop plants, especially maize, corn,
CC   barley, sorghum, soya, sugar cane, beet, tobacco or cotton plants.
CC   Overexpression of SGS2 can be used to increase resistance to viral
CC   infection in plants
XX
SQ   Sequence 6863 BP; 1862 A; 1354 C; 1485 G; 2162 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.45e-07       Length:         6863
Score:                  289.50         Matches:        190
Percent Similarity:     35.43%         Conservative:   120
Best Local Similarity:  21.71%         Mismatches:     297
Query Match:            3.91%          Indels:         269
DB:                     4              Gaps:           41

US-09-913-878A-2 (1-1402) x AAH77703 (1-6863)

Qy        535 PhePheGluIleLeuIleProSerProThrSerThrSerProSerValProProValVal 554
              |||         ::::::        ||||||:::|||   |||            |||
Db       2245 TTTGTTGCTCCTATTGTGAAGGATTTGACATCAAGTTCTTTCTCCCAGAAGACCTACGTT 2304

Qy        555 SerLysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSer 574
                 |||:::      ::::::                  ||||||   |||
Db       2305 TTTAAAAGAGTGAAGAGCATA------------------TTAACCGATGGG---TTTAAA 2343

Qy        575 LeuValGlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeu 594
              |||    ||||||::::::   :::|||    |||      |||
Db       2344 CTATGTGGTAGAAAATAC---AGTTTCTAGCATTCTCAGCCAAT--------------- 2385

Qy        595 ArgGluPheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhe 614
                       |||||||||                                      |||
Db       2386 ---------CAACTGAGA-----------------------GACCGCTCTGCATGGTTC 2412

Qy        615 PheAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValPro 634
              |||||||||   |||                        |||||||||
Db       2413 TTTGCTGAAGACGGG-----------------------AAAACACGT------------ 2436

Qy        635 AlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeu 654
                                                     ||||||   :::     |||:::
Db       2437 -------------------------------------GTGTCAGATATAAAAACATGGATG 2460

Qy        655 LeuGlnLeu---AspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGln 673
                 :::      |||   |||                |||      :::|||:::
Db       2461 GGGAAGTTCAAAGACAAGAATGTG-----------GCAAAATGTGCTGCTAGGATGGGC 2508

Qy        674 LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHis 693
              |||      |||   |||||||||   :::   :::   ||||||::::::
Db       2509 CTGTGCTTCTCCTCCACATATGCCACTGTAGATGTCATGCCTCACGAGGTT--------- 2559

Qy        694 LysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArg 713
                 |||::::::                  |||   |||   :::||||||:::|||
Db       2560 GACACTGAGGTTCCAGATATTGAGAGAAATGGGTATGTTTTCTCTGACGGAATTGGTACA 2619

Qy        714 MetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal------ 731
              ::::::   :::|||      :::   :::   |||   |||   ||||||
Db       2620 ATCACACCTGACCTCGCTGACGAAGTAATGGAGAAACTTAAGTTG---GATGTGCACTAC 2676

Qy        732 ---ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspVal 750
                 |||   |||   |||   |||:::         ||||||:::
Db       2677 AGCCCTTGTGCTTATCAGATACGTTACGCAGGTTTCAAAGGGGTT--------------- 2721

Qy        751 AspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln--------------- 765
                                      :::   :::||||||:::
Db       2722 -----------------------GTTGCTCGTTGGCCATCAAAAAGTGATGGAATCAGG 2757

Qy        766 ---------ArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluVal 782
                       |||                  |||   ||||||         ||||||:::
Db       2758 CTAGCCCTTCGAGACAGTATGAAGAAGTTCTTTTCCAAACAT---ACGATCTTGGAGATC 2814

Qy        783 ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeu 802
                 |||      :::      :::         ||||||   ||||::::::   :::|||
Db       2815 TGTTCC---TGGACGAGGTTTCAACCTGGGTTCTTAAATCGGCAGATAATTACCCTTCTA 2871
```

```
Qy        803 GluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsn 822
                                                       :::   |||    :::
Db       2872 TCC--------------------------GTACTAGGTGTTCCGGATGAAATATTCTGG 2904

Qy        823 AspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArg 842
              |||:::|||                         :::    ||||||||   ::::::
Db       2905 GATATGCAG---------GAATCCATGCTCTATAAACTGAACCGCATCCTTGATGATACA 2955

Qy        843 GlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValPro 862
                       :::|||    ::::::
Db       2956 GATGTGGCATTTGAAGTTCTCACGGCATCATGTGCT------------------------ 2991

Qy        863 PheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGly 882
                                      |||      |||        :::::::::::|||
Db       2992 ---------------------GAACAGGGAAACACTGCAGCTATCATGCTTAGTGCAGGT 3030

Qy        883 PheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLys 902
              |||   ||||||   :::   :::|||:::   :::        ::::::        :::
Db       3031 TTCAAACCAAAAACCGAGCCGCATCTACGCGGGATGTTGTCTTCAGTCAGAATTGCACAA 3090

Qy        903 CysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMet 922
                       |||:::   |||      |||   |||            ::::::
Db       3091 CTCTGGGGTCTCAGAGAAAAATCTCGTATTTTTGTTACTTCAGGAAGGTGGCTAATGGGT 3150

Qy        923 IleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer------ 940
                    |||      |||:::||||||      :::   :::     |||
Db       3151 TGCCTAGACGAAGCAGGGATACTTGAACATGGCCAATGCTTTATTCAAGTCTCTAAACCG 3210

Qy        941 -----------------------------SerLysPheArgAspGluGluGluSerPhe 950
                                           |||:::|||::::::   :::
Db       3211 TCTATAGAAAATTGTTTCTCCAAACATGGGTCTCGTTTTAAGGAGACAAAGACAGATCTG 3270

Qy        951 ThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAsp 970
              ::::::                   ||| :::|||:::::::::|||      |||   |||
Db       3271 GAAGTAGTTAAAGGCTAT---GTAGCCATTGCTAAGAATCCTTGTCTTCACCCAGGGGAT 3327

Qy        971 IleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIle 990
              ::::::   :::   ||||||      |||:::||||||   :::   |||   :::|||
Db       3328 GTAAGGATTTAGAAGCTGTTGATGTACCCCAGCTGCATCACATGTATGACTGCCTTATT 3387

Qy        991 PheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGly 1010
              |||      ||||||||   |||            :::   ||||||   |||   ||||||
Db       3388 TTCCCTCAGAAAGGTGATAGGCCGCATACAAACGAAGCTTCTGGCAGTGACCTTGACGGG 3447

Qy       1011 AspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMet 1030
              |||:::   :::|||   ||||||   :::::::::
Db       3448 GACCTGTACTTTGTGGCTTGGGATCAGAAACTCATC------------------------ 3483

Qy       1031 ProLeuGluProAspLeuSerArgTyr------LeuLysLysAspLysThrThrPheLys 1048
                    |||:::         |||      :::    |||              |||
Db       3484 ------CCTCCCAACAGGAAAAGCTATCCGGCCATGCATTATGATGCAGCTGAAGAGAAG 3537

Qy       1049 GlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyr------ 1066
                 |||                |||   |||      |||      |||
Db       3538 AGTTTA------------------GGCCGTGCTGTCAACCACCAGGTCAA-TATCTACTT 3578

Qy       1067 -------------------------------------AspMetIleGlnLysSerPhe 1073
                                                   |||   :::   |||   |||
Db       3579 CCCCACTTTTTTTCTGTTAGCTGAGATGATGACGCTTAAGATAAAATGGACAAGTGATTC 3638

Qy       1074 HisPhe--------------------AlaLeuGlnProAsnPheLeuGlyMetCysThr 1086
                 :::                      ::::::      ||||||      |||:::
Db       3639 CAGTATAACTCTCTGATATTTCTGGAAGTGGTGGAAACTTTGTTTTTGGATGCTTGTAGT 3698

Qy       1087 AsnTyr----------------------------------LysGluArgLeuCysTyr 1094
              :::|||                                  :::            |||
Db       3699 GATTACTTTTGCTTGCTTCTGAGTTCATAGATTGCGTGGCATAGGTTGTTAACAGCATAC 3758

Qy       1095 IleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeu 1114
                    :::   |||      |||      :::            :::      :::
Db       3759 TGTGTTCATCCTGTAGGAGCTAAG---ATGATGACATTTAAGAAAACTTTGTCATATATT 3815

Qy       1115 ValAspGlnSer------------------------------------------Lys 1119
              ||||||
Db       3816 GTTGATTCTGATTTGTATTTTTCGCATGCCTTACTGTGACTGTGACTGTTTAGTTACTAT 3875

Qy       1120 GlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu------------ 1135
              ||||||      :::   :::||||||:::   |||         |||
Db       3876 CAAGGGCAGTATCATGATGCTAGTAGCTTCTGCCAATGTAGTTATGAATTCTTTGTTATT 3935

Qy       1136 --------------------LeuLeuGlyGly---AlaLeuSerLeuProAspProMet 1147
                                  :::||||||      |||   |||   |||
Db       3936 TTTTCTTGCTGTGCTTGCCAAGTCTTGGGTAATTGATTGCTTTTGTTACAGGAC------ 3989

Qy       1148 TyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTyrLeuLys 1167
                                                       |||||||||:::
Db       3990 ---------------------------------------ATAATCGATTTC------ 4001

Qy       1168 PheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLys 1187
                    ||||||      :::::::::|||               |||
Db       4002 ------TTTGCAAGAAACTTGGCGAATGAGCAGTTGGGCACAATTTGCAAT--------- 4046

Qy       1188 AlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyr 1207
                                    ||||||               |||   :::   |||
Db       4047 ----------------------GCACACGTCGTTCATGCTGATAGAAGTGAGTATGGA 4082

Qy       1208 ThrPhePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeu 1227
                       :::|||                       ||||||         |||
Db       4083 GCCATGGACGAAGAA----------------------TGTTTGCTACTGGCAGAACTA 4118

Qy       1228 LysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuVal--------------- 1242
                       :::      |||      |||:::::::|||
Db       4119 GCTGCCACTGCAGTTGATTTCCCAAAGACA---GGGAAAATTGTGTCAATGCCCTTCCAC 4175

Qy       1243 ---LysAsnLysGluMetArgAsp------SerLysAspProTyrProVal--------- 1256
                 |||   |||         |||         |||:::   |||
Db       4176 CTAAAACCAAAACTCTACCCAGATTTCATGGGAAAAGAAGACTACCAAACTTACAAGTCG 4235

Qy       1257 -------------------------ArgValAsnGlnValTyrGluLysTrpCys--- 1266
                                       ||||||   :::||||||::::::
Db       4236 AACAAAATCTTGGGTCGGCTTTACAGACGGGTAAAAGAGGTTTATGATGAAGATGCAGAA 4295

Qy       1267 AlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysValIleArg 1286
              |||   :::   |||:::   |||   |||         ||||||:::
Db       4296 GCTTCCTCAGAAGAAAGCACAGACCCAAGTGCCATCCCTTATGACGCT------------ 4343

Qy       1287 LeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp 1301
              :::||||||:::                  |||         |||
Db       4344 GTTCTTGAAATACCGGGATTTGAAGATTTGATCCCTGAGGCATGG 4388
```

RESULT 10
AAC63742
ID   AAC63742 standard; cDNA; 3807 BP.
XX
AC   AAC63742;
XX
DT   07-FEB-2001 (first entry)
XX
DE   Soybean RNA-directed RNA polymerase FIS contig, SEQ ID NO: 11.
XX
KW   Soybean; plant; RNA-directed RNA polymerase; FIS; gene mapping;
KW   gene marker; plant virus resistance; plant breeding; ss.
XX
OS   Glycine max.
XX
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.

```
XX
PI   Odell JT,  Orozco EM,  Weng Z,  Sakai H,  Cahoon RE,  Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   P-PSDB; AAB28533.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 2; Page 49-50; 62pp; English.
XX
CC   The present sequence is one of a number of cDNA molecules which encode
CC   plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC   probes for genetically and physically mapping genes, and as markers for
CC   traits linked to those genes. They are useful for controlling gene
CC   expression and provide mechanisms to engineer plant virus resistance.
CC   They are also useful for plant breeding to develop lines with desired
CC   phenotypes
XX
SQ   Sequence 3807 BP; 1132 A; 662 C; 826 G; 1185 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.:               1.76e-07      Length:         3807
Score:                   283.00        Matches:        234
Percent Similarity:      32.43%        Conservative:   160
Best Local Similarity:   19.26%        Mismatches:     402
Query Match:             3.83%         Indels:         420
DB:                      3             Gaps:           53

US-09-913-878A-2 (1-1402) x AAC63742 (1-3807)

Qy    306 ProArgGluLysArgProValAspAlaThrValPheGluAlaGlyHisLeuIleGluSer 325
Db    135 CCACACGATCGTAGAGCTTCTGAT------ATTTTCGAAGAGGAAAGGCAAAAGAGTTA 188

Qy    326 ProSerLys---GlyArgThrThrLysSerHisIleAspAsnGlnProLeuSerSerSer 344
Db    189 GCATTTAGGATGGGAAAAACAATTGAGTTGTAT------GGATTCCCT---------ACA 233

Qy    345 SerGlnGlyGluThrSerPheSerThrTyrTyrGluSerPheProSerSerGly------ 362
Db    234 TCTGTGAATGTGTCTGATGTAAAGACATTTGTAGAGCAGTATACTGGTGAAGGAACTGTG 293

Qy    363 --------------------GlyGluGlyAlaIleProGluPro-------------- 370
Db    294 TTCGCCATTAAATTAAGACATGGAAAAGGTCGGGTTCCAAGAGCATTTGCAATTATTCAA 353

Qy    371 --------------------------------SerArgSerAsnGlyLeuAlaArgSer 379
Db    354 TTCACCACCGCAAATTCTGCTACATCTATGATGTCCAGAGCTAACAACATTTTGAGAACA 413

Qy    380 GluGluSerAlaArgSerGlnValGlnValHisAlaProValValAlaAlaArgLeuArg 399
Db    414 TTGCGGTATGGGACCTCCTATTTAAAAGCTCGG---------------GAAATGGAAAGA 458

Qy    400 AsnIleTrpProLysPheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGlu 419
Db    459 GATATTGTGCCAAGG------------------------------------------ 473

Qy    420 ValThrArgLeuPheMetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLys 439
Db    474 ---CCAAGGGTGTTTTTGCATAGTTTGGATGATGTGAAA---------------CTGTCT 515

Qy    440 TyrAspProSerTrpSerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArg 459
Db    516 TTTGGCTGTCAGATCTCAAAGGGAAGA---TTCTCTGTTTTATGGAAA------------ 560

Qy    460 LeuAspAlaPheArgGlyLysProPheProGluLysProProAsnAspValPheValThr 479
Db    561 ---------------------------------------AAGCAGGATGTTATTGTAAAT 581

Qy    480 AlaMetThrGly------------AsnPheGluSerLysGlySerAlaValValLeuSer 495
Db    582 TTTGGGAGTGGAATGAGAAAGATGCATTTCTTATTTCCCACAACAATGTGCAATACAAA 641

Qy    496 AlaValLeuAspTyr--------------------------AsnProAspAsnSerPro 506
Db    642 CTTGAGCTTTCATATGAGAACATTTGGAAGATTGAGCTGCATCGGCCACGGAATGAGACT 701

Qy    507 ThrAlaProLeuTyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeu 526
Db    702 ACACGT------TATCTG---------------------------------------TTG 716

Qy    527 ThrArgArgPheGlyProAspArgPhePheGluIleLeuIleProSerProThrSerThr 546
Db    717 ATTCAGTTACTTGGTGCTCCCCGGGTTTTTGAGAACGATGTA------CCTACATCAACA 770

Qy    547 SerProSerValProProValValSerLysGlnProGlyAlaValGluGluValIleGln 566
Db    771 AATATCTTTGATGATCCTTTGTTCAACTTCTTCAAAGATGCCCCTGATGAG------CAA 824

Qy    567 TrpLeu--------------------------ThrMetGlyGlnHisSerLeuVal--- 576
Db    825 TGGATCCGAGCAATTGATTTCACTCCAGAAAGTCGTATTGGGCAGTCCTCCGCCATATGT 884

Qy    577 ---------------GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArg 591
Db    885 CTGGAGCTTCCTAATGGCCGACAACTTCCAAATTTCAGGGAAAACTTTGCTTATTATGAG 944

Qy    592 LysProLeuArgGluPheGlnLeu------------------------------------ 599
Db    945 GAAAGTGAGAGGCAATACACTTTACAGACAGGAGTTCCCTTTTCTCAAAATTGGGGTCTT 1004

Qy    599 ------------------------------------------------------------ 599
Db   1005 GTCCCCATTGTTGCTCCTCCTCTAGGTGTTAAAATATCATATGACATCTTGTTTAAAGTC 1064

Qy    600 ---------------------------------------------------------Arg 600
Db   1065 AATTCATTGGTTCAACATGCATGTCTTGCAGGACCTGCACTTGATGGTGACTTCTATCGC 1124

Qy    601 AlaGluAspProLysProIleIleLysGluArgValHisPhePheAlaGlu--------- 617
Db   1125 TTGGTTGATCCACGTAGAATGCCCCGTGAATTTATTGAATATGCTTTAGAAAAGATTTAC 1184

Qy    618 ------ThrGlyIleThrPheArgPro---------------AspValPheLysThr--- 629
Db   1185 TATTCAAAGGAATTTTGTTATGAACCCACAAAGTGGCTGACTGATCAGTACAAAACATAC 1244

Qy    630 --------------------ArgSer-------------------------------- 631
Db   1245 CTTGAGTCAAAAAATCATCCTCGGTCACCTGCAATATCCTTGGATACAGGGTTGGTATAC 1304

Qy    632 ---------------ValValProAlaGlu---------------------------- 636
Db   1305 GTTCGCAGGGTTCAGATCACGCCTTGCAAAGTATACTTTTGTGGTCCAGAGATGAATGTC 1364

Qy    637 ---------------------------GluProValGluGln------ArgThrGluPhe 645
Db   1365 TCAAATCGTGTTCTCCGTCATTTCCGTGAACATATTGATAACTTTCTACGTGTTTCATTT 1424

Qy    646 LysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAsp------------------Asn 659
Db   1425 ---GTTGATGAAGAATTGGATAAACTGTTTTCAACTGATTTATCATCACGTTCACAGAAC 1481

Qy    660 AsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeu--------- 676
Db   1482 AAGAAAACTGAGATATACACCAGAATTCTTTCCATCCTTAAGAATGGCATAGTTGTTGGT 1541

Qy    677 SerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThrAsp 696
Db   1542 GATAAGAAGTTTGAATTTCTAGCATTCTCATCAAGTCAGTTGCGGGAAAACTCTCTCTGG 1601

Qy    697 LeuLeuSerProSerGlyThrGly------------------------------------ 704
```

```
Db   1602 ATGTTTGCTCCTACAGAAACTGGATGTACTGCTGCTTACATAAGGAAATGGATGGGAAAT 1661

Qy    704 ------------------------------------------------------------ 704

Db   1662 TTTAGCCAGATTAGGAATGTTGCTAAATATGCTGCTAGGCTGGGGCAATCTTTTGGTTCA 1721

Qy    704 ------------------------------------------------------------ 704

Db   1722 TCTACTGAAACTCTAAGTGTCCATAGGGATGAAGTTGAAATTATTCCTGATGTGAAGAAG 1781

Qy    705 --------------------GluValMetAsnAspGlyValGlyArgMetSerArgSer 717

Db   1782 CTTACATATGATGGAAACGAATATGTCTTCTCTGATGGAATTGGAAAAATATCTCTTGAA 1841

Qy    718 ValAlaLysArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGly 737

Db   1842 TTTGCCCAGAAAGTGGCTAAAAAATGTGGTTATGATTGCACTCCATCTGCCTTTCAGATT 1901

Qy    738 ArgPheGlySerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAsp 757

Db   1902 CGATATGGTGGGTACAAAGGAGTTGTGGCTGTTGACCCTAAATCATGCTACAAG------ 1955

Qy    758 TrpIleGluThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGln 777

Db   1956 ---TTATCACTGAGGAAGAGCATGCGGAAGTATGATTCAGAT---------------AAC 1997

Qy    778 ArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGln 797

Db   1998 ACAAAGTTAGATGTTTTGGCC---CGTAGTAAGTTTCAGCCATGTTATCTGAATCGGCAG 2054

Qy    798 LeuLeuProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGly 817

Db   2055 TTAATTTCTCTCTTATCC--------------------------ACTCTTGGTATCAAG 2087

Qy    818 AspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArg 837

Db   2088 GATGATGTTTTTGAGAAAAAACAAAGAGAAACTGTTAATCAACTGAACACTATACTAACA 2147

Qy    838 ProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSer 857

Db   2148 GATTCATTAAAGGCTCAGGAAGTTCTGGACTTAATGTCTGCT------------------ 2189

Qy    858 HisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPhe 877

Db   2190 ---GGAGAGATC-----------------------------ACTAATGTTCTGAAGGAG 2216

Qy    878 LeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAsp 897

Db   2217 ATGCTCATTTGTGGATACAAGCCTAATGAAGAACCATTCCTTTCAATGATGCTTCAAACA 2276

Qy    898 LeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSer 917

Db   2277 TTTAGGGCATCAAAACTTTTGGAATTGCGACTTAAATCTAGGATCTTTATTCCAAAAGGA 2336

Qy    918 AlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisVal 937

Db   2337 AGAGCAATGATGGGATGTCTAGATGAAACTAGAACCCTAGAATATGGTCAAGTATTTGTT 2396

Qy    938 GlyPheSerSer------------LysPheArgAspGluGluGluSerPheThrLeu--- 952

Db   2397 CAGTTCTCTAACAATAGGCTGCAGAATCTATCTGATGATTTTTTTCATATGATTTGCCA 2456

Qy    953 ---------LeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSer 969

Db   2457 AAGAATTATATGGTTAAAGGTAAGGTAGTTGTAGCAAAAAAYCCCTGCTTGCACCCAGGT 2516

Qy    970 AspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspValIle 989

Db   2517 GATGTGCGTGTTTTACAAGCTGTGGATGTGCCAGATTTGTACCACATGGTGGACTGTGTT 2576

Qy    990 IlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAsp 1009

Db   2577 GTTTTCCCTCAAAAAGGACCAAGACCTCATCCAAATGAGTGTTCGGGAAGTGATCTGGAT 2636

Qy   1010 GlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAlaGlu 1029

Db   2637 GGAGATATCTACTTTGTTTGTTGGGACCATGAATTGATT---------------CCTTCT 2681

Qy   1030 MetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGln 1049

Db   2682 CGCCCAATTGATCCA--------------------ATGGACTATACTGCTCCYGCAACT 2720

Qy   1050 LeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIle 1069

Db   2721 GTGGAATTGGATCAT-----------------------GATGTGATGATCGAGGAGGTT 2756

Qy   1070 GlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyr--- 1088

Db   2757 GAGGAGTATTTTGCCAATTACATAGTCAATGACAGTCTGGGAATAATTGCCAATGCACAC 2816

Qy   1089 ---------------LysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysPro 1103

Db   2817 ACTGTCTTTGCAGATAAAGAACATTTG---------------AAAGCAATGTCTGATCAA 2861

Qy   1104 AlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleVal 1123

Db   2862 TGTGTTAAGCTTGCAAGGTTGTTTTCAACAGCAGTTGACTTTCCTAAAACTGGTGTTCCA 2921

Qy   1124 PheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeu------------LeuGlyGly 1139

Db   2922 ------------------GCTGTTATACCTCCTGAACTTCATGTCAAAGAATATCCTGAC 2963

Qy   1140 AlaLeuSerLeuProAsp---ProMetTyrLysSerAspSerTrpLeuGlyArg------ 1156

Db   2964 TTCATGGAGAAGCCTGACAAACCCACATACAAATCGCATAACGTGATAGGAAAGCTCTTT 3023

Qy   1157 --------------------GlyGluProThrHisIleIleAspTyrLeuLysPheSer 1169

Db   3024 AGGGAAGTGAAAGAAATATCAACAAGTGCCGGCTCAATTACATCCTTCACAAAGTTG--- 3080

Qy   1170 IleAlaArgProAlaIleAspLysGluLeuGluAla------------------------ 1181

Db   3081 GTTGCGAGAGACTCTTACGACCATGAAATGGAAGTTGATGGCTTCATGGATTATGTTGAT 3140

Qy   1182 ---------PheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPhe 1198

Db   3141 GATGCTTTCTATCACAAAACCAATTATGACTACAAGTT-GGGAAATCTGATGGACTACTA 3199

Qy   1199 TrpAspProAspLeuAla---------------SerTyrTyrThrPhePheLysGluIle 1213

Db   3200 TGGGATCAAAACTGAAGCTGAAATCCTCGGTGGGAATATTATGAAAATGTCAAAATCTTT 3259

Qy   1214 SerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGly--- 1232

Db   3260 CAACAAAAGGAGGGATGCAGAAGC---------------------AATCAATATGGCTGT 3298

Qy   1233 GluValGluLysGluTyrGlyArg---LeuValLysAsnLysGlu 1246

Db   3299 GAGGTCCCTAAGGAAAGAGGCCAGGGCCTGGTTCAATGAAAACAG 3343
```

RESULT 11

```
AAC63739
ID   AAC63739 standard; cDNA; 3737 BP.
XX
AC   AAC63739;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Maize RNA-directed RNA polymerase EST sequence, SEQ ID NO: 5.
XX
KW   Maize; plant; RNA-directed RNA polymerase; expressed sequence tag; EST;
KW   gene mapping; gene marker; plant virus resistance; plant breeding; ss.
XX
OS   Zea mays.
XX
```

```
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT,  Orozco EM,  Weng Z,  Sakai H,  Cahoon RE,  Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   P-PSDB; AAB28530.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 2; Page 35-36; 62pp; English.
XX
CC   The present sequence is one of a number of cDNA molecules which encode
CC   plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC   probes for genetically and physically mapping genes, and as markers for
CC   traits linked to those genes. They are useful for controlling gene
CC   expression and provide mechanisms to engineer plant virus resistance.
CC   They are also useful for plant breeding to develop lines with desired
CC   phenotypes
XX
SQ   Sequence 3737 BP; 1063 A; 760 C; 863 G; 1051 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.47e-06      Length:         3737
Score:                  267.50        Matches:        156
Percent Similarity:     37.17%        Conservative:   112
Best Local Similarity:  21.64%        Mismatches:     296
Query Match:            3.62%         Indels:         157
DB:                     3             Gaps:           27

US-09-913-878A-2 (1-1402) x AAC63739 (1-3737)

Qy      645 PheLysValSerGlnMetLeuAsp---------Trp--------------------Leu 654
             |||    ||||||:::  |||         |||                    |||
Db     1472 TTTTCTTCAAGTCAGCTTCGAGATAACTCTGCATGGATGTTTGCTTCTCGGCAGGGATTG 1531

Qy      655 LeuGlnLeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGln--- 673
                     |||         |||         :::          :::
Db     1532 ACTGCGAGCGACATAAGGAAGTGGATGGGGGACTTTCGAGATATCAGAAATGTGGCAAAG 1591

Qy      674 ------------LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGln 689
                        ||||||   |||   :::   :::    |||||||:::    |||:::
Db     1592 TATGCTGCAATACTTGGGCAATCTTTCAGTTCCTCAACAGAAACTTTAAAAGTACACAAA 1651

Qy      690 IleArgHisHisLysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAsp 709
                        :::       ::::::   :::||||||      :::   :::|||
Db     1652 TCTGAGGTGGAACGAATTCCTGATATTACA---AATGGCACAAAGTACATATTCTCTGAT 1708

Qy      710 GlyValGlyArgMetSerArgSerValAla-------------------------Lys 720
            |||||||||:::::::|||   :::   |||                         |||
Db     1709 GGAGTTGGAAAGATCTCAGCTAATTTTGCAGTGGAGGTGGCTATGAAGTGCAAATTGAAA 1768

Qy      721 ArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGly 740
            |||                                   ||||||   |||   |||:::|||
Db     1769 CGCTTT-----------------------GCTCCTTCTGTTTTTCAGATAAGGTATGGC 1804

Qy      741 SerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGlu 760
                     ||||||:::      :::         ||||||                :::
Db     1805 GGTTACAAAGGTGTTGTCGCTGTA---------GATACAAGATCAAATCATAAGCTTTCT 1855

Qy      761 ThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeu 780
                     |||      |||::::::   :::   :::                  ||||||
Db     1856 TTGAGAAAAAGCATGTCAAAGTTCCAGTCAGAAAATATC---------------ACTCTT 1900

Qy      781 GluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuPro 800
            :::|||   :::      |||:::   :::         ||||||    ||||||:::
Db     1901 GATGTCCTTGCATAC---AGCAAGTACCAACCATGCTTCCTGAATCGGCAGTTGATTACT 1957

Qy      801 ValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeu 820
            :::|||                                     :::   |||   :::
Db     1958 CTTCTCTCA--------------------------ACACTTGGGGTTAGCGATAATGTC 1990

Qy      821 IleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGlu 840
                         |||::::::         |||         ||||||||   |||
Db     1991 TTTGAGCTAAAGCAGAAGGAAGCCTTAAGGCAG---------TTGAACAGAATGGTAACT 2041

Qy      841 PheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArg 860
                                           :::|||   |||   :::
Db     2042 ---------------------------GAACCACAGGCTGCTCGTGAAGCAGTTGAACTT 2074

Qy      861 ValProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsn 880
            :::|||                                 :::      |||::::::
Db     2075 ATGCCCATG---------------GGAGAGGTAACCAATGTAGTTAAAGAATTGTTGTCA 2119

Qy      881 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 900
               |||:::   |||      :::   ||||||      :::            :::
Db     2120 TGTGGCTACCAGCCTGATCATGAGCCATATCTTTCCATGCTGCTACAAACTTTTAGAGCA 2179

Qy      901 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 920
               |||         ||||||:::|||   |||   :::   :::          :::
Db     2180 TCCAAGCTTCTAGAGTTGAAAACAAAGTCAAGGATATTCATCACACAGGGGCGAGCAATG 2239

Qy      921 TyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSer 940
                     |||         |||:::      :::|||   |||         |||
Db     2240 ATGGGTTGCCTGGATGAAACCTGCACACTTAAGTACGGCCAGGTATTCGTCCAAGCTTCT 2299

Qy      941 SerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 960
                     |||   :::         ::::::               |||:::|||
Db     2300 TACAGTGCAGATGACCATCGCAAG---------GTCGTTGTAACTGGAAAAGTAGTTGTC 2350

Qy      961 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 980
            |||::::::|||         |||   |||||||   ||||||::::::   |||
Db     2351 GCCAAAAATCCTTGTCTCCACCCTGGTGACATACGGGTTCTCCAGGCTGTTGATGTTCCT 2410

Qy      981 GluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAla 1000
               ||||||   |||   |||   ::::::|||         :::|||      |||
Db     2411 GCTCTGCACCACTTGTTTGACTGTGTTGTCTTTCCACAGCAGGGACCAAGGCCGCACCCT 2470

Qy     1001 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 1020
               :::   ||||||   |||   |||||||||:::   :::|||   ||||||||
Db     2471 AATGAGTGTTCAGGGAGTGATCTTGATGGGGACATATATTTTGTTTCTTGGGATCCACAT 2530

Qy     1021 IleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeu 1040
            ::::::         :::         |||         |||      :::      |||
Db     2531 CTTATTCCAAGTCGTTTGGTGGATCCTATGGACTATACTCCAGCTTCAGCAGAAACATTA 2590

Qy     1041 LysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAla 1060
                  |||
Db     2591 GACCATGAT--------------------------------------------------- 2599

Qy     1061 LysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsn 1080
                     |||   :::   ||||||:::   |||            :::
Db     2600 ---------GTCACTATTGAGGAGATACAGGAGTACTTCACAAACTACATAGTTAATGAG 2650

Qy     1081 PheLeuGlyMetCysThrAsnTyr------------------LysGluArgLeuCysTyr 1094
               ||||||:::      |||                     :::||||||:::
Db     2651 AGTCTTGGGATTATCGCCAATGCGCATGTGGTCTTTACAGATCAGGAACGTATG------ 2704

Qy     1095 IleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeu 1114
                     :::   |||   :::   |||:::   |||
Db     2705 ---------AAAGCTGAGAGTCCACCGTGCGTTCAACTGGCCAAGCTCTTCTCTATAGCT 2755

Qy     1115 ValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArg 1134
             ||||||   |||   |||:::                       |||   :::
Db     2756 GTCGATTTCCCAAAGACTGGAGTGCCG-----------------GCTCTGATTCCACAT 2797
```

```
Qy       1135 GluLeuLeuGlyGlyAlaLeuSerLeuProAspProMet--------------------- 1147
Db       2798 GAGCTA------CATGTCAAGGAGTATCCTGACTTCATGGAGAAACTCGACAAAGTCACC 2851

Qy       1148 TyrLysSerAspSerTrpLeuGlyArg---------------GlyGluProThrHisIle 1162
Db       2852 TATGAATCAAAGGGTGTGATCGGGAAGCTCTATAGGGAAATAAAGAAGCACACACCACAC 2911

Qy       1163 IleAspTyrLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeu------Glu 1180
Db       2912 ATAAAGCACTTCACGAGGGAAGTGGCAAGGCGGTCTTATGACACCGATTTGATTGTTGAT 2971

Qy       1181 AlaPheHisAsnAlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAsp 1200
Db       2972 GGCTATGAAGATTACATTACTGAGGCTATAGAGTTCAAGGAAGAGTACGATTTCAGGCTG 3031

Qy       1201 ProAspLeuAlaSerTyrTyrThrPhePheLysGlu---------IleSerAspLysSer 1217
Db       3032 GGTAATCTTATGGACCACTATGGCATAAAAAGTGAAGCTGAGATAATAAGTGGATGTATT 3091

Qy       1218 ArgSerSerAlaLeuLeuPheThr------------------ThrLeuLysAsnArgIle 1231
Db       3092 CTAAAGATGGCAAAGAATTTCACCAAGAGTAGTGATGCTGATGCAATTAGAATGGCGGTG 3151

Qy       1232 GlyGluValGluLysGlu------------TyrGlyArgLeuValLysAsnLysGluMet 1247
Db       3152 AGATCTTTGAGGAAAGAAGCTAGGTCGTGGTTCAATGAGATGAGCACAGGAGAGGATGGC 3211

Qy       1248 ArgAspSerLysAspPro---------------TyrProValArgValAsnGlnValTyr 1262
Db       3212 CAAGATGCCATGGAGGCCAAGGCCTCTGCTTGGTACCATGTTACTTATCATCAGCAGTAC 3271

Qy       1263 ---------------------------------------------GluLysTrpCysAla 1267
Db       3272 TGGGGCAGCTACAATGAAGGGTATGATCGGCCGCATCTTATTAGCTTCCCATGGTGCGTA 3331

Qy       1268 IleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysValIleArgLeu 1287
Db       3332 TATGACAAGCTTGTGGCCATCAAGCAGGGGAGGAATCTCCTCACGCAGATGGATCGAAAC 3391

Qy       1288 Leu 1288
Db       3392 TTG 3394
```

```
RESULT 12
AAC63738
ID   AAC63738 standard; cDNA; 1281 BP.
XX
AC   AAC63738;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Maize RNA-directed RNA polymerase EST contig, SEQ ID NO: 3.
XX
KW   Maize; plant; RNA-directed RNA polymerase; expressed sequence tag; EST;
KW   gene mapping; gene marker; plant virus resistance; plant breeding; ss.
XX
OS   Zea mays.
XX
PN   WO200060097-A1.
XX
PD   12-OCT-2000.
XX
PF   06-APR-2000; 2000WO-US009105.
XX
PR   07-APR-1999;   99US-0128094P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Odell JT,  Orozco EM,  Weng Z,  Sakai H,  Cahoon RE,  Wang J;
PI   Zhong G;
XX
DR   WPI; 2000-679376/66.
DR   P-PSDB; AAB28529.
XX
PT   New nucleic acid fragments encoding RNA-directed RNA polymerase useful
PT   for controlling gene expression and providing mechanisms to engineer
PT   plant virus resistance.
XX
PS   Claim 2; Page 33; 62pp; English.
XX
CC   The present sequence is one of a number of cDNA molecules which encode
CC   plant RNA-directed RNA polymerase proteins. The sequences are useful as
CC   probes for genetically and physically mapping genes, and as markers for
CC   traits linked to those genes. They are useful for controlling gene
CC   expression and provide mechanisms to engineer plant virus resistance.
CC   They are also useful for plant breeding to develop lines with desired
CC   phenotypes
XX
SQ   Sequence 1281 BP; 359 A; 282 C; 299 G; 338 T; 0 U; 3 Other;

Alignment Scores:
Pred. No.:               3.21e-05      Length:          1281
Score:                   236.00        Matches:         84
Percent Similarity:      40.10%        Conservative:    74
Best Local Similarity:   21.32%        Mismatches:      158
Query Match:             3.19%         Indels:          78
DB:                      3             Gaps:            11

US-09-913-878A-2 (1-1402) x AAC63738 (1-1281)

Qy        698 LeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSerArgSer 717
Db         77 GTCACAACTGATGGCACTAAATACATATTTCAGATGGTATCGGGAAGATTTCTACTAGA 136

Qy        718 ValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal------ProSerAlaVal 735
Db        137 TTTGCCAGACAAGTCGCCAAATTAATTGGCTTAGACCCAGCTCATCCTCCTTCTGCTTTT 196

Qy        736 GlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAsp 755
Db        197 CAAATAAGGTATGGGGCTATAAAGGA------GTCATCACTATTGACCCTACATCCTTT 250

Qy        756 GluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAspPheValAspLys 775
Db        251 TTCAAT---CTTTCTCTGCGACCTAGTATGAAGAAGTTTGAATCG--------------- 292

Qy        776 HisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsn 795
Db        293 -----------------------------------------AAGAGCACTATGCTGAAC 310

Qy        796 Leu------------GlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLys 811
Db        311 ATTACAAATTGGAGTAAGTCTCAGCCATGTTATGTGAACCGTGAAATTATCTCTCTTCT 370

Qy        812 MetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGln 831
Db        371 TCAACATTGGGGATAAAGGATGAAGTATTTGANTCGATGCAACAAGATGACATGCACGAA 430

Qy        832 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 851
Db        431 TCAGATGGAATGCTAACAAATAAAGAA--------------------------------- 457

Qy        852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
Db        458 --------------------GCTGCTTTGTCTGTCCTAGGGAAAATTGGTGGCGGTGAT 496

Qy        872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeu 891
Db        497 ACCAAGACGGCAGCTGATATGCTTCTTCAAGGCTATGAACCAAGTTCAGAGCCTTACCTA 556

Qy        892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
Db        557 TTAATGATTCTTAAAGCCCATCGGGCTAATAGGCTGACCGACATAAGAACTCGGTGTAAG 616
```

```
Qy       912 IleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGlu 931
              |||  |||    :::             :::           ||          ||||||
Db       617 ATTCATGTCCAGAAAGGCCGTGTTCTTATTGGTTGTTTGGATGAAACTTGCAAATTAGAG 676

Qy       932 GluAsnGluValHisValGlyPheSerSerLysPheArgAspGlu---------GluGlu 948
                :::|||::::::          :::          :::::::::          |||:::
Db       677 TATGGCCAAGTTTACATCAGAATTACAAAGAATCGCAAGGAGCAGAAGTACAGTGAACAG 736

Qy       949 SerPhe-------------------ThrLeuLeuSerAspCysAspValLeuValAla 961
                |||                   |||   :::          |||   :::
Db       737 CCGTTCTTTTGCAACGATGATGGCAAAACAGCTGTAATTGTCGGAAAAGTTGCAATCACA 796

Qy       962 ArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGlu 981
             :::::|||          |||   |||::::::   :::    ||||||:::    |||
Db       797 AAAAACCCTTGTCTCCATCCTGGTGATGTCAGAGTACTTGAAGCTGTATATGACCCTGGA 856

Qy       982 Leu------HisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeu 999
             |||            |||   |||    ::::::|||       :::|||:::    |||
Db       857 TTGGATGCTAGGGGTCTTATTGATTGTGTTGTATTTCCTCAGAGAGGGGAAAGGCCTCAT 916

Qy      1000 AlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspPro 1019
                   :::    ||||||||||||   ||||||||||:::     ::::::   ||||||
Db       917 CCGAATGAATGCTCCGGGGGCGATTTGGATGGCGACCTCTTCTTTATTACTTGGGATGAC 976

Qy      1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
             :::::::::                |||:::|||   |||
Db       977 AAACTGATTCCGGAGAAGGTTGATGCACCTATG--------------------------- 1009

Qy      1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
                          |||   |||        :::       :::         |||
Db      1010 ---------GACTACACTGCAACGAGGCCACGCATAATGGACCATGCT----------- 1048

Qy      1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPhe 1073
                               |||   :::   |||||||||   |||
Db      1049 -----------GTTACACTTGAGGAAATTCAGAAGCACTTC 1078
```

RESULT 13
ACA42620
ID ACA42620 standard; DNA; 4551 BP.
XX
AC ACA42620;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #24277.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU38750.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 30490; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4551 BP; 779 A; 1608 C; 1519 G; 645 T; 0 U; 0 Other;

Alignment Scores:			
Pred. No.:	0.04	Length:	4551
Score:	195.50	Matches:	271
Percent Similarity:	32.95%	Conservative:	189
Best Local Similarity:	19.41%	Mismatches:	477
Query Match:	2.64%	Indels:	465
DB:	7	Gaps:	64

US-09-913-878A-2 (1-1402) x ACA42620 (1-4551)

```
Qy        86 AlaSerSerAsnTrp-ValProLysAlaHisAlaAspProAspThrLeuProTrpSerLy 105
             ::::::|||   |||  ||||||:::|||        :::|||   |||      |||   ::
Db       257 TCGAACAGCCGCTGGAGTTCCTCAAGCGGCAGAAGAACCTGC-ACCGGGTCCTGGCCGAG 315

Qy       105 sGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeuGl 125
              ::::       ||||||             |||    :::
Db       316 GGACAGTTCAAGGGCGAAGGCCTGTCGAACACCCTGGCCTATCTCT------------- 361

Qy       125 uValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArgTh 145
                                                        |||   |||        ::
Db       362 -------------------------------------GGGACAACTGGTATCGCAAG 381

Qy       145 rLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAspGl 165
              :   :::|||                      |||||||||             :::
Db       382 TCGCGCCGGGTGATGCAGCGCGCCTTCTCCTACGAGTCGCGGCAACAGGTCACCGAGGCG 441

Qy       165 uProAlaAsnValThrPheAlaAspPro---ProLysArg-SerLeuThrArgSerAlaT 184
                :::          :::|||   |||   |||:::|||  :::|||   |||:::|||
Db       442 GTGCCCGAGCTGAAGATGGGCAACGCCATGACCGCGCCGACGCCCTACGACGCGCAGCCG 501

Qy       184 hr------------GlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspP 200
                           |||      |||        :::||||||
Db       502 CAATACGCGCCGATTCCCGGCCTGCCGCTGGTCAGCCAGTTCCGCTA---------TCCG 552
```

```
Qy    200 roValAsnThrGlySerLysArgProSerLeuGluSerGluAsn---------LeuAsnG 217
Db    553 CTGTGGGACGCCAAGCCCATCGAGCCGCCGCAGGACGTCAAGCTGGCCGGTTCCTCCAGC 612
Qy    217 lnCysThrLysArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProP 237
Db    613 GAGTTCATCAACGTCGTGCCGGGCAATGTCTACATCCCGCTGGGCAAGCTGAAGCCGGGC 672
Qy    237 roValProIleAlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgA 257
Db    673 CTGTACCTGGTCGAGGCGCTTGTCGGCAAGTACCG---CGCCACCACCGTGGTGTTCGTC 729
Qy    257 spProThrAlaThrGlyHisArg---ArgAlaAspGlnValAspSerPheAspThrSerG 276
Db    730 TCCAACAGCGTGGCGGTGAGCAAGGTCGCCGGCGACGAGCTGCTGGTCTGGACCGCACGC 789
Qy    276 lnGlyThrSerTyrGlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrG 296
Db    790 AA----------------------------GCACGAAGGCACCCCGGTGCCCGACACC 819
Qy    296 lnSerSerPheGluAlaProProSer--------------GlnProArgGluLysArgP 311
Db    820 AAGGTGCTGTGGACCGACGGCCTGGGCGTGATGAGCAGCGGCAATACCGACGCCGACGGC 879
Qy    311 roValAspAlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgT 331
Db    880 CTGCTGCGGCTCAAGCACGCCAGTCCGGAGCGCTCCTACGT------GATCGGCG-AGGA 932
Qy    331 hrThrLysSerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerP 351
Db    933 CCGCGAGG-----------------------GCGGGGTCTTCGTCTCCGAGAACTTCTA 968
Qy    351 heSerThrTyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluPro- 370
Db    969 TTACGACAGCGAGATCTACGACACCAAGATCT--------------ACGCCTTCACCGA 1013
Qy    371 --SerArgSerAsnGlyLeuAlaArg-----------------SerGluGluSerAlaA 384
Db   1014 CCGCCCGCTGTACCGGCCGGGCGACTGGGTCTCGTTGAAGATGGTCGGCCGCGAGTTCAA 1073
Qy    384 rgSerGlnValGlnValHisAlaPro-----------------------------ValV 394
Db   1074 GGACGCCCGGCAATCCCAGGCCGCCGCCAGCGCGCCGGTGCGCCTGAGCGTGATCGACGC 1133
Qy    394 alAlaAlaArgLeuArgAsnIleTrpProLysPheProLysTrpLeuHisGluAlaProL 414
Db   1134 CAGCGGCACGGTATTGCAAAGCCTGGACCTCC---GCTTCGACGCGAAGAGCGGCGCCAA 1190
Qy    414 euAlaValAlaTrpGluValThrArgLeuPheMetHisCysLysValAspLeuGluAspG 434
Db   1191 TGGCCGCTTCC-----------------------AACTGCCGGAG-AACGCCGTCGCCG 1225
Qy    434 luSerLeuGlyLeuLysTyrAsp---------ProSerTrpSerThrAlaArgAspValT 451
Db   1226 GCGGCTACGAGTTGCGCTTCGACTACCGCGGGCAGACCTACAGCAGCGCCTTCCGCGTCG 1285
Qy    451 hrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArgGlyLysPro---------- 467
Db   1286 CCGAATACATCAAGCCGCACTTCGAAGTCGCCCTCGACCTGGCCAAGCCGGACTTCAAGA 1345
Qy    468 ----------------------------------PheProGluLysProPro------- 473
Db   1346 CCGCCGAGCCGGTGAAGGGCGAGATCGTCCTGCTCTATCCGGACGGCAAGCCGGTGGCCA 1405
Qy    474 -------------------------------------------------AsnAspValP 477
Db   1406 ACGCGCGCCTGCAACTGAGCCTGCGCGCCCAGCAGTTGTCGATGGTGGACAACGAACTG- 1464
Qy    477 heValThrAlaMetThrGlyAsnPheGluSerLysGlySerAlaValValLeuSer---- 495
Db   1465 --------CAGTATCTCGGCCAGTTCCCGGTGGAACTGAGCAGCACCGAACTGACCACCG 1516
Qy    496 -------------AlaValLeuAspTyrAsnProAspAsnSerProThr---------- 507
Db   1517 ACGGCAAGGGTCGCGCCGCCATCGAATTGCCGCCGGCCGAGAAACCCAGTCGCTACATGC 1576
Qy    508 -------------------AlaProLeuTyrLeuValLysLeu---LysProLeuMetP 520
Db   1577 TGACCATCTTCGCCAGCGACGGCGCCGCCTACCGGGTCAAGACCAGCAAGGAGATCCTCA 1636
Qy    520 heGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhe------------- 535
Db   1637 TCGAGCGCGGCGCCGCGCGCTACCGCCTGAGCGCGCCGCAGCGCTTCAGCGCCGCCGGCG 1696
Qy    536 -----------PheGluIleLeuIleProSerProThrSerThrSerProSerValProP 552
Db   1697 AGAAGGTCGAGTTCAGCTATGCCAGCGAACAGCCGACCCCGCTCAAGCCGAGC------- 1749
Qy    552 roValValSerLysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyG 572
Db   1750 ----------------------------------AGCTACCAGTGGATCCGCCTGGAGG 1774
Qy    572 lnHisSer--------------LeuValGlyArgGlnTrpArgAlaPhePheAlaLysA 587
Db   1775 ACCGCGCCACCGACAGCGGTCCGGTCGCCGACGGCCGCTTCGCGCTGACCTTCGAGCGCC 1834
Qy    587 spAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysProI 607
Db   1835 CCGGTACCTATTCCGTGGAGCTGCGC-----------------GACGACAAGGGCCAGC 1876
Qy    607 leIleLysGluArgValHisPhePheAlaGluThrGlyIleThrPheArgProAsp---- 625
Db   1877 TCCTTGGCGCTACCGGCCACTCGGTGAGCGGCGAGGGCGTGAAGTCGGTGCCGGGCACCG 1936
Qy    626 --------ValPhe--------------------------LysThrArgSerValValP 634
Db   1937 TCGAGGTGGTGTTCGACAAGCCCGAGTACCGCACCGGCGAGGAAGCCTCCGCACTGATCA 1996
Qy    634 roAlaGluGluProValGlu-----------------GlnArgThrGluPheLysValS 648
Db   1997 CCTTCCCCGAGCCGGTGGAGGATGCGCTGTTGTCCCTGGAGCGCGACAAGGTCGAGGCCA 2056
Qy    648 erGlnMetLeu-----------AspTrpLeu---LeuGlnLeuAspAsnAsnThr---- 661
Db   2057 CCGCCCTGCTGTCCAAGGGCGCCGACTGGCTGCGCCTGGAAAAGCTCAATCCGACCCAGT 2116
Qy    662 --------TrpGlnProHisLeuLysLeuPheSer------------------------ 670
Db   2117 ACCGCGTGTGGATCCCGGTGCGCGAGGAGTTCTCGCCGAACCTGACCTTCTCGGTGCTCT 2176
Qy    671 ----------------------------------ArgIleGlnLeuGlyLeuSerLysT 679
Db   2177 ACACCAAGGGCGGCGACTACAGCTTCCAGAACGCCGGGATCAAGGTCGGCATGCCCCAGG 2236
Qy    679 hrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThr---------- 695
Db   2237 TAGAGATCGACATCGCCACCGACAAGGAGCGCTACGAACCGGGCGAGACGGTGACCGTGA 2296
Qy    696 ----------------------AspLeuLeuSerProSerGlyThrGlyGluValMetA 708
Db   2297 CCCTGGCCACGCGCTTCGCCGGCAAACCGGTTTCCAGCCACCTGACCGTCAGCGTGGTGG 2356
Qy    708 snAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyL 728
Db   2357 ATGAGATGGTCTATGCCCTGCAGGCGGAGATCGCCCC-CGGCATCGACCAGTTCTT---- 2411
Qy    728 euGlyAspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValI 748
Db   2412 --------CTACCATCCGCGGCGCAACAACGTGC-GCACCAGCG--------------- 2446
Qy    748 leAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln---ArgL 767
Db   2447 ----------------------------------CCAGCCTGGCCTTCATCAGCTACGA 2471
Qy    767 ysTrpGluCys--------------------AspPheValAspLysHisGlnArgThrL 780
```

```
Db       2472 CGTGGCGTTGCCCGGCAGCACCAGCGCACCCGGACGG-GCGAACCGCAGCGAGCGCGGCG 2530

Qy        780 euGluVal------ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnL 798
              :::::|||       |||          ::::::    :::|||            :::|
Db       2531 TCAAGGTACTGGAGCGGCCGCGCCGCGAGGACGTCGACACCGCCGCCTGGCAGCCGGAGC 2590

Qy        798 euLeuProValLeuGluAspArgAlaArgAspLysValLysMet---------------A 813
              ||:::          :::    :::|||          :::|||                |
Db       2591 TGGTCACCGACGCCCAGGGCAAGGCCAGCTTCAGCTTCCGCATGCCCGACTCGCTGACCC 2650

Qy        813 rgGlnAlaIleGlyAspArgLeuIleAsnAsp---------------LeuGlnArgGlnP 828
              ||       |||       |||    |||:::|||                :::::::|||
Db       2651 GCTGGCGGATCACCGCGCGGGCCATCGACGACAACGGCCAGGTCGGCCAGAAGAAGCAGT 2710

Qy        828 heSerGluGlnLysHisAlaLeu------------AsnArgProValGluPheArgGln- 843
              ||          :::       |||                :::    |||        |||||||||
Db       2711 TCCTTCGCTCGGAGAAGCCGCTGTACCTGAAGTGGAGCGGGCCGACTCGCTTCCGCCAGG 2770

Qy        844 ----------------------TrpValTyrGluSerTyrSerSerArgAlaThrArgV 856
                                    :::|||:::
Db       2771 GTGACCAGCCCGACCTCGGCCTGTTCGTCTTC--------------------------- 2802

Qy        856 alSerHisGlyArgValPro------------PheLeuAlaGlyLeuProAspSerGln- 871
                :::    |||    |||                |||:::|||    |||    |||||
Db       2803 --AACCAGGGCGAACAGCCGGTCAAGGCCGAGCTGCTCAGCGGGCCGCCTGGCAGCCAGC 2860

Qy        872 --GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrL 891
                :::||||||                      :::    |||       |||:::    |
Db       2861 GCAGCCAGACCCTGGAGCTGGCCAAGGGAGTCAACTACATTCCGCTGGCGCAACAGCCGC 2920

Qy        891 euGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuA 911
              ||    |||       |||       :::    |||:::          ::::::    :::|||
Db       2921 TCAGCGATGGCGACTGGAGCGCCGAGTTGCGCCAGGACGGGCAGGTCCGCGACCGCCTGG 2980

Qy        911 snIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuG 931
                :::|||                      ::::::|||||    |||    |||      :
Db       2981 CGGTGCGCTTCAAC------------------CTGCTGGCCGACGGCTGGCAGGTGGAGC 3022

Qy        931 luGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheT 951
              ::
Db       3023 AG---------------------------------------------------------- 3024

Qy        951 hrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspI 971
                   :::    :::       :::       |||    :::|||          |||:::|||
Db       3025 -----ATGCAGAACCTCAGCCTGGCGGCGGCGAGCAACCCGCTGCAACTGCCGGCGGACG 3079

Qy        971 leGlnArgValArg-----------------------AlaValPheLysProGluLeuH 983
                :::    ||||||                         |||    :::          |||
Db       3080 CGCGCGACGTACGCCTGCGCCTGGCCGATGGCCCGGCCGCTGCCTAC---------CTGG 3130

Qy        983 isSerLeuLysAspValIleIlePheSerThrLysGlyAsp------------------ 996
                :::|||    |||::::::    :::       |||
Db       3131 GCAATCTCGATGATCTGCTGGAGTATCCCTACGGCGGCGTCGAGCAGACCGCCAGCCAAC 3190

Qy        997 --ValProLeuAla---------LysLysLeuSerGlyGlyAsp--------------- 1007
                :::||||||:::                 |||:::||||||:::
Db       3191 TGCTGCCGCTGAGCATCGCCTATCCGGCGCTGGCCGGCGGCGAGCCGCGGATTCGCGACC 3250

Qy       1008 --------------------------------------------TyrAspGlyAspMetA 1013
                                                          |||          |
Db       3251 GCCTGCGCCTGATCATGCAGAACAGTCGCCTGCGCCTGGTGCAGATGGCCGGCCCGGACG 3310

Qy       1013 laTrpValCysTrpAspProGluIleValAspGly------PheValAsnAla------- 1028
              |||||    |||          ||||||||                |||:::    |||
Db       3311 CCTGGTTCGCCTGGTGGGGTGGCGACGTCGATGGCGACGCTTTCCTCACCGCCTACGCCT 3370

Qy       1029 ---------------------------------------GluMetProLeuGluProAspL 1036
                                                     ::::::|||    |||
Db       3371 ACTACGCCGACTGGTACGCCAGCCGCGCACTGGAGATCCAACTGCCGGCCGAA------- 3423

Qy       1036 euSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyT 1056
                                                  ::::::::::::::        :::
Db       3424 ------------------------------CACTGGCAGCGCATCCTCGAGCCCTAC---- 3450

Qy       1056 hrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheA 1076
                              ||||||:::    |||          ::::::|||::::::        |
Db       3451 -----------GCCAAGCAGGCGACGCAGACCCCGCTGCTGCAGCGTGCGCTGATCCTGG 3499

Qy       1076 laLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleA 1096
              ||
Db       3500 CTTTC----------------------------------------------------- 3504

Qy       1096 snAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValA 1116
                             :::          :::    :::::::::|||:::|||    |||::::
Db       3505 -----------GCCCGCGACATGCAATTGCCGGTGAACACCCTGCTCGGCGGCCTGCTCA 3553

Qy       1116 spGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluL 1136
              ::                                      |||       |||    |||    |||
Db       3554 ACGAT-------------------CTGGCCAACGCCGGCGAAGGGCAGGCGCGCGCGGAGC 3595

Qy       1136 euLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyA 1156
                |||                                  ::::::|||       |||
Db       3596 CGCTG--------------------------------GAGGCCGACGACGGCCTGGTGC 3622

Qy       1156 rgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleA 1176
                |||:::|||          :::                :::          :::
Db       3623 TGGGCGACCCGGACTCGGCCGTCGGCCTGGCCGCGGCGCGAGTGCTGGCGGTGGACCTTG 3682

Qy       1176 spLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAsp------------ 1191
                ::::::|||             |||       ||||||:::
Db       3683 CACGGCAACTGCGGGTCGCGGTGCCCGCGCCGCTGGCCGCCCAGGCCGAGACGGCGACGC 3742

Qy       1192 --------ThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrP 1209
                          |||    |||       |||    |||          ||||||||
Db       3743 AGCGCCTGCGTGAGGCCGGCTTGCCGTTCACCGATGCCTTGCTGGCCTCT---------- 3792

Qy       1209 hePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeu 1223
                   :::          |||       ::::::||||||||||||
Db       3793 -----CGCTCGGCCGTGGATGGCCAGCAGGCCAGCGCTCTGCTG 3831

RESULT 14
AAV05287
ID   AAV05287 standard; DNA; 49377 BP.
XX
AC   AAV05287;
XX
DT   17-OCT-2003  (revised)
DT   21-MAY-1998  (first entry)
XX
DE   The soraphen biosynthesis gene cluster from Sorangium cellulosum.
XX
KW   Polyketide synthase; PKS; biosynthesis; soraphen; SorR; SorA; SorB; SorM;
KW   biosynthetic module; beta-ketoacylsynthase; acyltransferase;
KW   ketoreductase; beta-ketone processing domain; cytostatic agent;
KW   antimicrobial agent; phytopathogenic fungi; transgenic plant;
KW   biological control; ss.
XX
OS   Polyangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             383..760
FT                   /*tag= a
FT                   /product= "SorR"
FT                   /note= "gene product highly homologous to the reductase
FT                   domains of type I PKSs such as eryA from
FT                   Saccharopolyspora erythraea"
FT   CDS             927..19874
FT                   /*tag= b
FT                   /product= "SorA"
FT                   /note= "gene product is highly homologous to type I PKSs
FT                   that are known to be involved in the synthesis of
FT                   polyketide compounds"
FT   misc_feature    942..7115
FT                   /*tag= c
FT                   /note= "module 1 of SorA"
FT   misc_feature    7203..12884
```

```
FT                   /*tag=  d
FT                   /note= "module 2 of SorA"
FT   misc_feature    13455. .19616
FT                   /*tag=  e
FT                   /note= "module 3 of SorA"
FT   misc_feature    19870. .24556
FT                   /*tag=  h
FT                   /note= "module 1 of SorB"
FT   misc_feature    19871. .46318
FT                   /*tag=  f
FT   CDS             19871. .46318
FT                   /*tag=  g
FT                   /product= "SorB"
FT                   /note= "gene product is highly homologous to type I PKS
FT                   genes"
FT   misc_feature    24638. .30820
FT                   /*tag=  i
FT                   /note= "module 2 of SorB"
FT   misc_feature    30881. .35446
FT                   /*tag=  j
FT                   /note= "module 3 of SorB"
FT   misc_feature    35528. .40114
FT                   /*tag=  k
FT                   /note= "module 4 of SorB"
FT   misc_feature    40190. .46318
FT                   /*tag=  l
FT                   /note= "module 5 of SorB"
FT   CDS             46851. .47891
FT                   /*tag=  m
FT                   /product= "SorM"
FT                   /note= "gene product is homologous to the
FT                   methyltransferase from Streptomyces hygroscopicus that is
FT                   involved in the synthesis of the polyketide rappamicin"
XX
PN   US5716849-A.
XX
PD   10-FEB-1998.
XX
PF   14-DEC-1996;   96US-00764233.
XX
PR   24-AUG-1993;   93WO-US007954.
PR   08-JUN-1994;   94US-00258261.
PR   09-OCT-1996;   96US-00729214.
XX
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Ligon JM,  Hill DS,  Neff S,  Beck JJ,  Ryals JA,  Schupp T;
XX
DR   WPI; 1998-158369/14.
XX
PT   DNA encoding Sorangium cellulosum polypeptide(s) - used for, e.g.
PT   biosynthesis of soraphen useful as antimicrobial agent against
PT   phytopathogenic fungi.
XX
PS   Claim 2; Col 47-90; 64pp; English.
XX
CC   The present sequence contains a cluster of genes that encode polyketide
CC   synthases (PKSs) that are involved in the synthesis of soraphens in
CC   Sorangium cellulosum. The proteins encoded by the present sequence are
CC   SorR, SorA, SorB and SorM. SorA and SorB contain biosynthetic modules
CC   which contain a beta-ketoacylsynthase, an acyltransferase, a
CC   ketoreductase and an acyl carrier protein domain, as well as beta-ketone
CC   processing domains. S. cellulosum soraphens are useful as a cytostatic
CC   and antimicrobial agent active against phytopathogenic fungi. Soraphen-
CC   producing transgenic plants or biological control agents can also be
CC   produced, which may reduce crop losses and nutritional deprivation for
CC   local populations in many parts of the world. (Updated on 17-OCT-2003 to
CC   standardise OS field)
XX
SQ   Sequence 49377 BP; 7247 A; 19522 C; 14477 G; 8131 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:          1.13         Length:       49377
```

```
Score:                    192.00        Matches:        225
Percent Similarity:       31.27%        Conservative:   109
Best Local Similarity:    21.07%        Mismatches:     339
Query Match:              2.60%         Indels:         395
DB:                       2             Gaps:           51

US-09-913-878A-2 (1-1402) x AAV05287 (1-49377)

Qy          5 ThrProArgLysArgAsnSerPro-----------------------------ValGlu 14
                :::||||||    :::    ||||||                               :::
Db      10696 TCGCCTCGGTGCAAGCGCTCGCCACGCGCATCACGTCCGCCGAGCAGCTCCGCACCCCAG 10755

Qy         15 GluIleIleAsnArgLeuAsnAsnAspTyrAsnLeuGlyLeuGlnCysValAlaAspThr 34
              |||:::         :::   :::     :::   |||         |||  |||
Db      10756 GAGCTTCCCACCACGATGCCCTCTTCCGCGTCGACTGGAGCGAGCTGCCGAGCC------ 10809

Qy         35 ThrLeuThrProHisArgArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHis 54
                    |||||||||   :::      |||   |||
Db      10810 ------CCACCTCACCGTCTGGAGCCCCAAGCGCCGTCCTTCTCGGCATCGGCGGCCTCG 10863

Qy         55 AspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArgLysAspAspSerLeuAsnGln 74
                         |||   |||:::                        |||      :::
Db      10864 ACCTCGCCCCCGAGGTGCCTCTCGCCCGCGTCGCCGACCTCGCTGCCCTCCAGAGCGCGC 10923

Qy         75 AlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSerSerAsnTrpValProLysAla 94
              :::               ::::::|||   ::::::|||   :::|||:::
Db      10924 TCGACCAAGGCGCTTCGCCTCCAGGCCTCGTCGTCGTCCCCTTCATGGCTA--------- 10974

Qy         95 HisAlaAspProAspThrLeu--------------------------------------- 101
                    :::|||   |||
Db      10975 ------GAACCGCCGACGACCTCATCCAGAGCGCCCACTCCATCACCGCGCGCGCCCTCG 11028

Qy        102 --------------------ProTrpSerLysGluProProArgAlaAlaThrAlaGly 114
                                  |||   :::      ||||||||||||::::::
Db      11029 CCCTGCTGCAAGCCTGGCTGGCCGACGAACGCCTCGCCTCCTCGCGCCTCGTCCTGCTCA 11088

Qy        115 GlnGlnTrpAlaLeuGlnThrValLeuLeuGluValLeuAsnArgPheMetProProPro 134
                                                              :::|||   |||
Db      11089 CCCGAC-------------------------------GCGCCATCGCTGCCCGCGCCG 11115

Qy        135 AsnAsnThrProGlyArgThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArg 154
                   |||      ||||||            |||||||||   |||        |||
Db      11116 ATGAAGACG---TCAAGGACCTCGCTCACGCCCCTCTCTGGGGGCTCGCACGCTCCGCGC 11172

Qy        155 ProThrSerThrAsnThrLysArgLysAspGluProAlaAsnValThrPheAlaAspPro 174
                   :::|||         |||         :::      |||         |||
Db      11173 AAAGCGAACACCCAGAACTCCCGCTCTTTCTCG---TCGACCTGGACCTCAGTGAGGCCT 11229

Qy        175 Pro-------------------------LysArgSerLeuThrArg---SerAlaThr 184
              |||                         :::   |||:::||||||   |||||||||
Db      11230 CCCAGCACACCCTGCTCGCCGCGCTCGAAACAGGAGAGCGTCACTCGCGTCTCCGCAACG 11289

Qy        185 GlyProPro-----------------IleHis-------------------------- 189
                   |||               :::|||
Db      11290 GAAAACCCTTCATCCCGAGATTGGCGAATGCACGCTCGAAGGATGAGCTCATCGCCCCGG 11349

Qy        190 ------------GlyAlaAlaIleProLeuLysPhe---ProAspProValAsnThrGly 204
                          ||||||:::|||         |||   |||         :::|||
Db      11350 ACGCGTCCAACTGGCGCCTCCATA---------TTCCGACCAAAGGCAACTTCGACGCGC 11400

Qy        205 SerLys-----------------------ArgProSerLeuGluSerGluAsnLeuAsn 216
              |||                          |||||||||        ::::::::   :::
Db      11401 TCACCCTCGTCGACGCCCCTCTAGCCCGTGCGCCCCTCGCACACGGCCAAGTCCGCGTCG 11460

Qy        217 GlnCysThrLysArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaPro 236
                 ||||||
Db      11461 CCGTGCACG--------------------------------------------------- 11469

Qy        237 ProValProIleAlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArg 256
              |||   |||      ||||||:::            |||               ||||||
Db      11470 CCGCAGCCTTCAATTTCCGCGATGTCCTCGACACCCTTGGTCTGTATCCGGGCGACGCGG 11529

Qy        257 AspProThrAlaThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGln 276
```

```
Db   11530 GACCGCTCGGCG---GCGAAGGCGCAGGCATCGTTACTGAAGTCGGTCCAGGTGTTTCGC 11586

Qy     277 GlyThrSer---------------TyrGlySerSerValPheSerAlaCysArgHisAsn 291
Db   11587 GGTACACCGTAGGCGACCGGGTGATGGGGATCTTCGGCGCAGCTTGCGGTCCCACGGCCA 11646

Qy     292 GlnSerThrThrGlnSerSerPheGluAlaProProSerGlnProArgGluLysArgPro 311
Db   11647 TCGCCGACGCCCGCATGATCTGCCCCATCCCCCACGCCTGGTCCTTCGCCCAAGCCGCCA 11706

Qy     312 ValAspAlaThrValPheGluAlaGlyHisLeuIleGluSerProSer------LysGly 329
Db   11707 GCG-------------------------------------TCCCCATCATCTATCTCACCG 11730

Qy     330 ArgThrThrLysSerHisIleAspAsn------GlnProLeuSerSerSerSer------ 345
Db   11731 CCTACTACGGACTCGTCGATCTCGGGCATCTGAAACCCAATCAACGTGTCCTCATCCATG 11790

Qy     346 ---------------GlnGlyGluThrSerPheSerThrTyrTyrGluSerPheProSer 360
Db   11791 CGGCCGCCGGCGGCGTCGGGACGGCCGCCGTTCAGCTCGCACGCCACCTCGGCGCCGAGG 11850

Qy     361 SerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLeu------AlaArg 378
Db   11851 TCT---------------TTGCCACCGCCAGCGCAGGGAAGTGGAGCGCTCTCCGCGCGC 11895

Qy     379 SerGluGluSerAlaArgSerGlnValGlnValHisAla-ProValValAlaAlaArgLe 398
Db   11896 TCGGCTTCGACGACGCGCACCTCG---CGTCCTCACGTGACCTGGACTTCGAGCAGCACT 11952

Qy     398 uArgAsnIleTrpProLysPheProLysTrpLeuHisGlu-------------------- 411
Db   11953 T------------CCTGCGCTCCACGCATGGGCGCGGCGTGGATGTCGTCCTCGACTGCT 12000

Qy     412 ----------------------------AlaProLeuAlaValAlaTrpGluValThrAr 422
Db   12001 TGGCACGCGAGTTCGTCGACGCTTCGCTGCGCCTCATGCCGAGCGGTGGACGCTTCGTCG 12060

Qy     422 g-LeuPheMetHisCysLysValAspLeu---GluAspGluSerLeuGlyLeuLysTyrA 441
Db   12061 AGATG---------GGCAAGACGGACATCCGTGAGCCCGACGCGGTCGGCGTCGCCTACC 12111

Qy     441 spProSerTrpSerThrAlaArgAspValThrAspIle---------------------- 453
Db   12112 CTGGTGTCGTTTACCGCGCCTTCGACCTCATAGAGGCCGGACCGGATCGAATAGAGCAGA 12171

Qy     453 ------------------------------------------------------------ 453
Db   12172 TGCTCGCAGAGCTGCTCAGCCTCTTCGAGCGCGGTGCGCTTCGTCCGCCGCCCATCACAT 12231

Qy     454 --TrpLysThrLeuTyrArgLeuAspAlaPheArg-----------------------G 465
Db   12232 CTTGGGACATCCGTCATGCCCCCCAGGCCTTTCGCGCGCTCGCTCAGGCGCGGCATGTTG 12291

Qy     465 lyLys---------------ProPheProGluLysProProAsnAspValPheValThrA 480
Db   12292 GGAAGTTCGTCCTCACCATACCCCGTCCCATAGACCCCGAAGGCACCGTCCTCATCACGG 12351

Qy     480 laMetThrGly------------------------------------------------- 483
Db   12352 GAGGCACCGGCACGCTAGGAGCCCTGGTCGCACGCCATCTCGTCGCAAGACACGGCGCCA 12411

Qy     483 ------------------------------------------------------------ 483
Db   12412 AGCACCTGCTTCTCACGTCGAGGCAGGGCGCGCACGCTCCGGGCGCCGAGGCCTCGCGAA 12471

Qy     484 --AsnPheGluSerLysGlySerAlaValValLeuSerAlaValLeuAspTyrAsnPro- 502
Db   12472 CCGAGCTCGAAGCGCTGGGGGCCTCTGTCACACTTCGCGCGTGCGACGCGGCCGACCCAC 12531

Qy     503 --------------------AspAsnSerProThrAla---ProLeuTyrLeuValLysL 515
Db   12532 GCGCCCTCCAAGCCCTCTTGGACTCCATCCCGAGCGCTCACCCGCTCACCGCCGTCGTCC 12591

Qy     515 euLysProLeuMetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgP 535
Db   12592 ACGCCGCGGGCGCCCTCGACGACGGC------CTGCTCGGCGCCATGAGCCCCGAGCGCA 12645

Qy     535 hePheGluIle-LeuIlePro--------------------------SerProThrSer 545
Db   12646 TCGACCGCGTCTTTGCCCCCAAGCTCGATGCTGCTTGGCACTTGCATGAGCTCACCCAAG 12705

Qy     546 ThrSerProSerValProProValValSerLys--------------------------- 556
Db   12706 ACAAGCCCCTCGCCGCCTTCGTCCTCTTCTCGTCCGCTGCTGGCGTCCTTGGTAGTCCAG 12765

Qy     557 ------------------------GlnProGlyAlaValGluGluValIleGlnTrpLeu 568
Db   12766 GTCAGTCGAACTACGCCGCGGCCAATGCCTTCCTCGATGCGCTCGCGCATCACCGGCGTG 12825

Qy     569 ThrMetGly---------------GlnHis-SerLeuValGlyArg-------------- 578
Db   12826 CCCACGGGCTCCCGGCCTCCTCGCTCGCATGGGGCTATTGGGCCGAGCGCAGTCGAATGA 12885

Qy     579 ----------GlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLys-------- 592
Db   12886 CCGAGCACCTCAGCGCCGCCGATGTTTCTCGCATGAGGCGCGCCGGCGTCCGGCCCCTCG 12945

Qy     593 -------------------ProLeuArgGluPheGlnLeuArgAlaGlu---------As 603
Db   12946 CCACAGACGAGGCGCTCTCCCTCTTCGATGCGGCTCTCTTGCGGCCCGAGCCCGCCCTGG 13005

Qy     603 pProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIleThrPheAr 623
Db   13006 TCCCCGCACGCTTCGACGTGAACGCGCT------------------------------CG 13035

Qy     623 gProAspValPheLysThrArgSerValValProAlaGluGluProValGluGlnArgTh 643
Db   13036 GCGCGAATGCCGACGAGGTGCCCCCGCTGTTCCAGCGTCTCGTCCGCGCTCGCGTCGCAC 13095

Qy     643 rGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAspAsnAsnThrTrpGl 663
Db   13096 GCAA--------------------------------------GGCCGCCAGCAATACCG 13116

Qy     663 nProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTyrAlaIleMe 683
Db   13117 CCCTGGCCTCTTCACTCTCTCAGCGCCTCTCCTCCCTCCCGCCCGCAGAAAGCGAGCGCT 13176

Qy     683 t------ThrLeuGluProHisGlnIleArgHisHisLysThrAspLeuLeu-------- 698
Db   13177 TCCTCCTCGATCTCGTCCGCACCGAAGCCGCCACCGTCCTCGGCCTCGCCTCATTCGAAT 13236

Qy     699 -------SerProSerGlyThr-GlyGluValMetAsnAspGlyVal------------- 711
Db   13237 CGCTCGATCCCCATCGCCCTCTCCAAGAGCTCGGCCTCGATTCTCTTATCGCTCTCGAGC 13296

Qy     712 -----GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyA 730
Db   13297 TCCGAGGTCGACTCGCCGCGGCCACCGGGCTGCGACTCCAACCTACTCTC---CTCTTCG 13353

Qy     730 spValProSer---AlaValGlnGlyArgPhe--------------------------- 739
Db   13354 ACTATCCAACCCCGGCTGCACTCTCACGCTTTTTCACGACGCAGTTCTTCGGGGAAACCA 13413

Qy     740 ------------------------------GlySer------------------------ 741
Db   13414 CCGACCGTCCCGCAGCGCCGCTCACCCCGGCGGGAAGCGAAGACCCTATCGCCATCGTGT 13473

Qy     742 -----------------------------------AlaLysGlyMetTrpValIleAspV 750
Db   13474 CGATGAGCTGCCGCTTCCCTGGCGACGTGCGCACGCCCGAGGATCTCTGGAAGCTCTTGC 13533

Qy     750 alAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluC 770
Db   13534 TCGAT---------GGGAAAGATGCCATCTCCAGCTTTCCCCAGAATCGCGGTTGGAGTC 13584
```

```
Qy        770 ysAspPheValAsp 774
                |||   :::|||
Db      13585 TCGATGCGCTCGAC 13598

RESULT 15
AAZ90112
ID   AAZ90112 standard; cDNA; 4148 BP.
XX
AC   AAZ90112;
XX
DT   19-MAY-2000  (first entry)
XX
DE   Human antizuai-2 (AZ-2) gene.
XX
KW   Antizuai-2; AZ-2; human; breast cancer; tumour suppressor; ss;
KW   malignancy progression marker; malignancy reversion; chromosome 10q26.
XX
OS   Homo sapiens.
XX
PN   WO200000503-A1.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US014482.
XX
PR   26-JUN-1998;   98US-0090747P.
XX
PA   (CHEN/) CHEN H.
PA   (BISS/) BISSELL M.
XX
PI   Chen H,  Bissell M;
XX
DR   WPI; 2000-170903/15.
DR   P-PSDB; AAY78795.
XX
PT   New nucleic acid encoding a tumor suppressor or marker, used for
PT   diagnosis, monitoring progress or treatment, and gene therapy of breast
PT   cancer.
XX
PS   Claim 2; Fig 5; 120pp; English.
XX
CC   This sequence represents the human antizuai-2 (AZ-2) gene sequence. AZ-2
CC   is a variant of the antizuai-1 (AZ-1) gene (see AAZ90111). The AZ-1 gene
CC   is located on chromosome 10q26, and encodes a protein that acts as a
CC   tumour suppressor or marker of malignancy progression or reversion. The
CC   AZ-1 protein and its variants are tumour suppressors, AZ-1 interacts with
CC   E-cadherin and beta-catenin. Detecting low levels of AZ-1 nucleotide or
CC   amino acid sequences are used to diagnose a breast cell malignancy, also
CC   for monitoring disease progression, particularly assessment of
CC   therapeutic efficacy. The nucleotide sequence is used in in vivo or ex
CC   vivo gene therapy, and AZ-1 polypeptides are used for treating or
CC   preventing breast cancer. AZ-1 polypeptides are also used to raise
CC   specific antibodies, for diagnostic detection of AZ-1. Fragments of the
CC   AZ-1 nucleotide sequence are useful as probes or primers for detecting
CC   expression of the AZ-1 gene
XX
SQ   Sequence 4148 BP; 1107 A; 1140 C; 1082 G; 819 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0.0944        Length:          4148
Score:                  188.50        Matches:         197
Percent Similarity:     33.95%        Conservative:    154
Best Local Similarity:  19.05%        Mismatches:      386
Query Match:            2.55%         Indels:          298
DB:                     3             Gaps:            45

US-09-913-878A-2 (1-1402) x AAZ90112 (1-4148)

Qy          1 MetAsnProIleThrProArgLysArgAsnSerProValGluGluIleIleAsnArgLeu 20
                |||   |||      |||   |||:::::::::|||      :::::::::::::::
Db        676 ATGTTACCTTCGGTTCCT---AAGAAGGATGCTCCA------AGAGTCATGGATAAAGTC 726

Qy         21 AsnAsnAspTyrAsnLeuGlyLeuGln---------------CysValAlaAspThrThr 35
                :::|||         |||   :::                  ||||||||
Db        727 ACTTCAGATGAGACCAGAGGTGCGGAAGGAACAGAAAGTTCACCTGTGGCAGATGATATC 786

Qy         36 LeuThrProHisArgArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHisAsp 55
              :::   |||         :::|||
Db        787 ATCCAGCCCGCTGCCCCTGCAGACCTG---------------------------GAAAGC 819

Qy         56 LysIleTyrArgAlaLeuAsnPheLeuTyrTrpArgLysAspAspSerLeuAsnGlnAla 75
                        |||   ::::::                  |||   :::   |||
Db        820 CCAACCTTAGCTGCCTCTTCCTAC------------CACAGTGATGTTGTTGGCCAGGTC 867

Qy         76 GluAlaAsnPhePheIleGluAlaLysAlaAlaSerSerAsnTrpValProLysAlaHis 95
                    :::      :::   :::   |||                     :::|||
Db        868 TCTACGGATCTGATAGCCCAGAGGAGTTCCGATTCTGAA------------GAGGCATTT 915

Qy         96 AlaAspProAspThrLeuProTrpSerLysGluProProArgAlaAlaThrAlaGlyGln 115
                 |||::::::            |||   ||||||
Db        916 GAGACCCCGGAGTCAACGACCCCTGTCAAAGCTCCGCCAGCTCCACCC------------ 963

Qy        116 GlnTrpAlaLeuGlnThrValLeuLeuGluValLeuAsnArgPheMetProProProAsn 135
                                                              |||||||||
Db        964 -------------------------------------------------CCACCACCC--- 972

Qy        136 AsnThrProGlyArgThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArgPro 155
                                                       |||    :::     |||
Db        973 ---------------------------------------CCCGAAGTCATCCCAGAACCC 993

Qy        156 ThrSerThrAsnThrLysArgLysAspGluProAla------------AsnValThrPhe 171
                 :::                :::||||||                   |||
Db        994 GAGGTCAGCACACAGCCACCCCCGGAAGAACCAGGATGTGGTTCTGAGACAGTCCCTGTC 1053

Qy        172 AlaAspProProLysArg---SerLeuThrArgSerAlaThrGlyProProIleHisGly 190
                |||   |||:::      |||:::         |||         ||||||   |||
Db       1054 CCTGATGGCCCACGGAGCGACTCGGTGGAAGGAAGTCCCTTCCGTCCCCCGTCACACCCC 1113

Qy        191 AlaAlaIleProLeuLysPheProAspProValAsnThrGlySerLysArgProSerLeu 210
                :::                  |||:::         |||         :::|||
Db       1114 TTCTCTGCCGTCTTCGATGAAGACCAGCCGATA---GCCAGCAGTGGGACTTACAACTTG 1170

Qy        211 GluSerGluAsnLeuAsn------------------------------------GlnCys 218
              :::   :::|||:::
Db       1171 GACTTTGACAACATTGAGCTTGTGGATACCTTTCAGACCTTGGAGCCTCGTGCCTCAGAC 1230

Qy        219 ThrLysArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProVal 238
                 |||      :::||||||:::         |||         :::         |||
Db       1231 GCTAAGAATCAGGAGGGCAAAGTG------AACACACGGAGGAAGTCCACGGATTCCGTC 1284

Qy        239 ProIleAla---SerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAsp 257
              ||||||:::   |||   |||      :::   :::   :::   ::::::
Db       1285 CCCATCTCTAAGTCTACACTGTCCCGGTCGCTCAGCCTGCAAGCCAGTGACTTTGATGGT 1344

Qy        258 ProThrAlaThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGly 277
                 :::::::::|||:::      |||      :::               |||   |||
Db       1345 GCTTCTTCCTCAGGCAATCCCGAGGCCGTGGCCCTTGCCCCAGATGCATATAGCACGGGT 1404

Qy        278 ThrSerTyrGlySerSerValPheSerAlaCysArg---------------------His 290
              :::|||      ||||||               :::
Db       1405 TCCAGCAGTGCTTCTAGTACCCTTAAGCGAACTAAAAAACCGAGGCCGCCTTCCTTAAAA 1464

Qy        291 AsnGlnSerThrThrGlnSerSerPheGluAlaProProSerGlnProArgGluLysArg 310
                 :::   ||||||:::         |||   ||||||   :::      ::::::
Db       1465 AAGAAACAGACCACCAAGAAACCCACAGAGACCCCCCCAGTGAAGGAGACGCAACAGGAG 1524

Qy        311 ProValAspAlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArg 330
              |||   :::   ::::::      :::|||      :::   |||   :::|||
Db       1525 CCAGATGAAGAGAGCCTTGTCCCCAGTGGGGAGAATCTAGCATCTGAGACGAAAACGGAA 1584

Qy        331 ThrThrLysSerHis------------------IleAspAsnGlnProLeuSerSerSer 344
              :::   |||:::                        ::::::   ||||||
Db       1585 TCTGCCAAGACGGAAGGTCCTAGCCCAGCCTTATTGGAGGAGACGCCCCTTGAG------ 1638

Qy        345 SerGlnGlyGluThrSerPheSerThrTyrTyrGluSerPheProSerSerGlyGlyGlu 364
```

```
Db       1639 -----------------------------------------CCCGCTGTGGGGCCCAAA 1656

Qy        365 GlyAlaIleProGluProSerArgSerAsnGlyLeuAlaArgSerGluGluSerAlaArg 384

Db       1657 GCTGCCTGCCCTCTGGACTCAGAGAGTGCAGAAGGGGTTGTCCCCCCGGCTTCTGGAGGT 1716

Qy        385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys 404

Db       1717 GGCAGAGTGCAGAACTCACCCCCTGTCGGGAGGAAAACGCTG------------------ 1758

Qy        405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424

Db       1759 -----------------------CCTCTTACCACGGCCCCGGAGGCAGGGGAGGTAACC 1794

Qy        425 MetHisCysLysValAspLeuGluAspGlu---------------SerLeuGlyLeuLys 439

Db       1795 CCATCGGATAGCGGGGGGCAAGAGGACTCTCCAGCCAAAGGGCTCTCCGTAAGGCTGGAG 1854

Qy        440 TyrAsp---------------ProSerTrpSerThrAlaArgAsp--------------- 449

Db       1855 TTTGACTATTCTGAGGACAAGAGTAGTTGGGACAACCAGCAGGAAAACCCCCCTCCTACC 1914

Qy        450 -----------------------------------------ValThrAspIleTrpLys 455

Db       1915 AAAAAGATAGGCAAAAAGCCAGTTGCCAAAATGCCCCTGAGGAGGCCAAAGATGAAAAAG 1974

Qy        456 ThrLeuTyrArgLeuAspAlaPheArgGlyLysProPheProGluLysPro------Pro 473

Db       1975 ACACCCGAGAAACTTGACAACACTCCTGCCTCACCT---CCCAGATCCCCTGCTGAACCC 2031

Qy        474 AsnAspValPheValThrAlaMetThrGlyAsnPheGluSerLysGlySerAlaValVal 493

Db       2032 AATGACATCCCCATTGCTAAAGGTACTTACACCTTTGAT--------------------- 2070

Qy        494 LeuSerAlaValLeuAspTyr------AsnProAspAsnSerProThrAlaProLeuTyr 511

Db       2071 ------------ATTGACAAGTGGGATGACCCCAATTTTAACCCTTTTTCTTCCACCTCA 2118

Qy        512 LeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrArgArgPheGly 531

Db       2119 AAAATGCAGGAGTCTCCCAAACTGCCCCAACAA------------TCATACAACTTTGAC 2166

Qy        532 ProAspArgPhePheGluIleLeuIlePro------------SerProThrSerThrSer 547

Db       2167 CCAGACACCTGTGATGAGTCCGTTGACCCCTTTAAGACATCCTCTAAGACCCCCAGCTCA 2226

Qy        548 ProSerValProProValValSerLysGlnProGlyAlaValGluGluVal-IleGlnTr 567

Db       2227 CCTTCTAAATCCCCAGCCTCCTTTGAGATCCCGGCCAGTGCTATGGAAGCCAATGGAGTG 2286

Qy        567 pLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPhe---PheAlaLy 586

Db       2287 GACGGGGATGGGCTAAACAAGCCCGCCAAGAAGAAGAAGACGCCC-CTAAAGACGGTGAA 2345

Qy        586 sAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspPro----- 604

Db       2346 AAAGTCGCCAAAACGGTCTCCTCTCTCTGAT------CCACCTTCCCAGGACCCCACCCC 2399

Qy        605 ------------------LysProIleIleLysGluArgValHisPhePheAlaGluTh 618

Db       2400 AGCTGCTACACCAGAAACACCACCAGTGATCTCTGCGGTGGTCCACGCCACAGATGAGGA 2459

Qy        618 rGlyIle--------------------------------------------------- 620

Db       2460 AAAGCTGGCGGTCACCAACCAGAAGTGGACGTGCATGACAGTGGACCTAGAGGCTGACAA 2519

Qy        621 ------------------------------ThrPheArgProAspValPheLysThrAr 630

Db       2520 ACAGGACTACCCGCAGCCCTCGGACCTGTCCACCTTT---------GTAAACGAGACCAA 2570

Qy        630 gSerValValProAlaGluGluProValGluGlnArgThrGluPheLysValSerGlnMe 650

Db       2571 ATTCAGTTCACCCACTGAGGAG---TTGGATTACAGAAACTCCTATGAAATTGAATATAT 2627

Qy        650 t----------------------------------------LeuAspTrpLeuLeuGl 656

Db       2628 GGAGAAAATTGGCTCCTCCTTACCTCAGGACGACGATGCCCCGAAGAAGCAGGCCTTGTA 2687

Qy        656 nLeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLe 676

Db       2688 CCTTATGTTTGACACTTCTCAGGAGAGCCCTGTCAAGTCATCTCCCGTCCGCATGTCAGA 2747

Qy        676 uSer-----------------------------LysThrTyrAlaIleMetThr----- 684

Db       2748 GTCCCCGACGCCGTGTTCAGGGTCAAGTTTGAAGAGACTGAAGCCCTTGTGAACACTGC 2807

Qy        685 ------------------------------LeuGluProHisGlnIleArgHisHisLy 694

Db       2808 TGCGAAAAACCAGCATCCTGTCCCACGAGGACTGGCCCCTAACCAAGAGTCACAC----- 2862

Qy        694 sThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMe 714

Db       2863 -------TTGCAGGTGCCAGAGAAATCCTCCCAGAAGGAGCTGGAGGCCATGGGCTTGGG 2915

Qy        714 tSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal-------- 731

Db       2916 CACCCCTTCAGAAGCGATTGAAATTAGAGAGGCTGCTCACCCAACAGACGTCTCCATCTC 2975

Qy        732 -ProSerAlaValGlnGlyArgPheGlySerAlaLys--------------GlyMetTr 746

Db       2976 CAAAACAGCCTTGTACTCCCGCATCGGGACCGCTGAGGTGGAGAAACCTGCAGGCCTTCT 3035

Qy        746 pVal------IleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSe 764

Db       3036 GTTCCAGCAGCCCGACCTGGACTCTGCCCTCCAGATCGCCAGAGCAGAGATCATAACCAA 3095

Qy        764 rGlnArg---------LysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGl 781

Db       3096 GGAGAGAGAGGTCTCAGAATGGAAAGATAAATATGAAGAAAGCAGGCGGGAAGTGATGGA 3155

Qy        781 uValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProVa 801

Db       3156 AATGAGGAAAATAGTGGCCGAGTATGAGAAGACC---------------ATCGCTCAGAT 3200

Qy        801 lLeuGluAspArgAlaArgAspLys----------------------------------Va 810

Db       3201 GATAGAGGACGAACAGAGAGAGAAGTCAGTCTCCCACCAGACGGTGCAGCAGCTGGTTCT 3260

Qy        810 lLysMetArgGlnAlaIleGlyAsp---------------ArgLeuIleAsnAspLeuGl 825

Db       3261 GGAGAAGGAGCAAGCCCTGGCCGACCTGAACTCCGTGGAGAAGTCTCTGGCCGACCTCTT 3320

Qy        825 nArgGlnPheSerGluGlnLysHisAlaLeu---------AsnArgProValGluPheAr 842

Db       3321 CAGAAGATATGAGAAGATGAAGGAGGTCCTAGAAGGCTTCCGCAAGAATGAAGAGGTGTT 3380

Qy        842 gGlnTrpValTyrGluSerTyrSerSerArgAlaThrArg 855

Db       3381 GAAGAGATGTGCGCAGGAGTACCTGTCCCGGGTGAAGAAG 3420
```

Search completed: March 31, 2004, 07:25:55
Job time : 1026.39 secs

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         March 31, 2004, 06:31:56 ; Search time 161.141 Seconds
                                           (without alignments)
                                           4828.317 Million cell updates/sec

Title:          US-09-913-878A-2
Perfect score:  7397
Sequence:       1 MNPITPRKRNSPVEEIINRL..........YEVLGDDDFDGIGFTGNGDY 1402

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:      1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09913878/runat_29032004_082009_8395/app_query.fasta_1.2254
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878_@CGN_1_1_82_@runat_29032004_082009_8395 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	297.5	4.0	3731	3	US-08-811-583-1	Sequence 1, Appli
	2	192	2.6	49377	1	US-08-764-233A-1	Sequence 1, Appli
	3	182	2.5	4026	4	US-09-252-991A-10214	Sequence 10214, A
	4	178.5	2.4	28958	1	US-08-258-261B-6	Sequence 6, Appli
	5	178.5	2.4	28958	1	US-08-456-837-6	Sequence 6, Appli
	6	178.5	2.4	28958	1	US-08-457-342-6	Sequence 6, Appli
	7	178.5	2.4	28958	1	US-08-457-646A-6	Sequence 6, Appli
	8	178.5	2.4	28958	1	US-08-458-076A-6	Sequence 6, Appli
	9	178.5	2.4	28958	1	US-08-764-233A-4	Sequence 4, Appli
	10	178.5	2.4	28958	1	US-08-457-335A-6	Sequence 6, Appli
	11	178.5	2.4	28958	1	US-08-729-214-6	Sequence 6, Appli
	12	178.5	2.4	28958	3	US-09-028-934-6	Sequence 6, Appli
	13	174	2.4	14770	4	US-09-220-132-30	Sequence 30, Appl
c	14	171	2.3	52297	3	US-09-426-436-1	Sequence 1, Appli
c	15	171	2.3	52297	4	US-08-705-557-1	Sequence 1, Appli
	16	162	2.2	15872	3	US-09-105-537-1	Sequence 1, Appli
	17	162	2.2	15872	4	US-09-091-609-1	Sequence 1, Appli
	18	162	2.2	15872	4	US-09-091-609-3	Sequence 3, Appli
	19	160.5	2.2	2943	1	US-08-042-747A-7	Sequence 7, Appli
c	20	160	2.2	8312	4	US-09-620-312D-1048	Sequence 1048, Ap
	21	158	2.1	2419	4	US-09-624-693A-12	Sequence 12, Appl
c	22	158	2.1	77536	4	US-09-410-551B-1	Sequence 1, Appli
	23	158	2.1	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	24	156.5	2.1	2472	4	US-09-252-991A-886	Sequence 886, App
c	25	156.5	2.1	2955	4	US-09-252-991A-797	Sequence 797, App
	26	156	2.1	13987	2	US-08-804-227C-13	Sequence 13, Appl
	27	156	2.1	44377	2	US-08-804-227C-7	Sequence 7, Appli
	28	156	2.1	44377	2	US-08-804-198-1	Sequence 1, Appli
	29	156	2.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	30	155.5	2.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	31	155.5	2.1	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	32	155	2.1	11907	3	US-08-061-376-4	Sequence 4, Appli
	33	154.5	2.1	77536	4	US-09-410-551B-1	Sequence 1, Appli
c	34	154	2.1	30001	1	US-08-125-468-1	Sequence 1, Appli
c	35	154	2.1	30001	2	US-08-474-933-1	Sequence 1, Appli
c	36	154	2.1	33529	3	US-09-144-085-3	Sequence 3, Appli
	37	153.5	2.1	11283	2	US-08-603-753D-3	Sequence 3, Appli
	38	153.5	2.1	11283	3	US-09-099-753-3	Sequence 3, Appli
	39	153.5	2.1	11283	3	US-08-986-106-3	Sequence 3, Appli
c	40	153.5	2.1	13613	3	US-09-105-537-3	Sequence 3, Appli
	41	152.5	2.1	2941	3	US-09-425-383-1	Sequence 1, Appli
c	42	152.5	2.1	36519	3	US-08-923-137-2	Sequence 2, Appli
	43	151.5	2.0	11385	2	US-08-639-501-1	Sequence 1, Appli
	44	151.5	2.0	11385	3	US-09-044-946-1	Sequence 1, Appli
	45	151.5	2.0	11385	3	US-09-044-908-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-811-583-1
; Sequence 1, Application US/08811583
; Patent No. 6218142
;  GENERAL INFORMATION:
;    APPLICANT:  Wassenegger, Michael
;    APPLICANT:  Riedel, Leonhard
;    APPLICANT:  Schiebel, Winfried
;    APPLICANT:  Sanger, Heinz
;    TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING
;    TITLE OF INVENTION:  POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
;    TITLE OF INVENTION:  RNA-DIRECTED RNA POLYMERASE (RdRP)
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & NEAVE
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/811,583
;      FILING DATE:  05-MAR-1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley, James F.
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  MPG-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-596-9000
;      TELEFAX:  212-596-9090
```

```
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3731 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Tomato
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  194..3535
US-08-811-583-1

Alignment Scores:
Pred. No.:               3.73e-16     Length:         3731
Score:                   297.50       Matches:        189
Percent Similarity:      35.78%       Conservative:   137
Best Local Similarity:   20.75%       Mismatches:     339
Query Match:             4.02%        Indels:         247
DB:                      3            Gaps:           35

US-09-913-878A-2 (1-1402) x US-08-811-583-1 (1-3731)

Qy     562 GluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpArg 581
Db    1349 GAAGACATAGATAACTTTCTTCGTGTT---------TCTTTTGTTGATGAGGAGTGGGAG 1399

Qy     582 AlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAla 601
Db    1400 AAACTGTATTCTACAGACTTA--------------------------------------- 1420

Qy     602 GluAspProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIleThr 621
Db    1421 ---TTACCAAAA------------------------------GCAAGTACTGGAAGTGGT 1447

Qy     622 PheArgProAspValPheLys------------------ThrArgSerValValProAla 635
Db    1448 GTCAGGACAAACATCTATGAGAGGATCTTATCAACTCTGCGGAAAGGCTTTGTAATTGGT 1507

Qy     636 GluGluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeu 655
Db    1508 GATAAAAAATTTGAA---TTTCTTGCATTTTCATCGAGCCAGTTG--------------- 1549

Qy     656 GlnLeuAspAsnAsnThrTrp------------GlnProHisLeuLysLeu--------- 668
Db    1550 ---CGGGATAATTCAGTGTGGATGTTTGCATCAAGACCTGGCCTTACTGCAAATGATATA 1606

Qy     669 ------------------PheSerArgIle------------------------GlnLeu 674
Db    1607 AGAGCTTGGATGGGTGATTTTTCGCAGATCAAGAATGTCGCAAAATATGCTGCCAGACTT 1666

Qy     675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLys 694
Db    1667 GGTCAATCTTTTGGTTCCTCCAGAGAGACTTTG---------AGTGTTCTTAGGCATGAG 1717

Qy     695 ThrAspLeuLeu---------SerProSerGlyThrGlyGluValMetAsnAspGlyVal 711
Db    1718 ATTGAAGTTATTCCCGATGTAAAGGTTCATGGAACCAGCTATGTCTTTTCTGATGGAATT 1777

Qy     712 GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspVal 731
Db    1778 GGTAAAATATCTGGTGACTTTGCTCATAGAGTTGCCTCAAAATGTGGCCTTCAATATACC 1837

Qy     732 ProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAsp 751
Db    1838 CCATCTGCTTTCCAGATTCGTTATGGTGGATATAAAGGTGTTGTGGGTGTTGATCCGGAT 1897

Qy     752 AspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluCysAsp 771
Db    1898 TCATCAATGAAG---------TTGTCTTTGAGAAAGAGCATGTCGAAATATGAATCAGAC 1948

Qy     772 PheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysSer 791
Db    1949 AACATA---------------AAGTTAGATGTCCTTGGA---TGGAGCAAATATCAGCCT 1990

Qy     792 AlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLysValLys 811
Db    1991 TGTTATCTTAATCGTCAACTGATTACGCTCTTGTCT------------------------ 2026

Qy     812 MetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGln 831
Db    2027 ---ACACTTGGAGTGAAAGATGAAGTTCTCGAACAGAAGCAAAAGGAAGCTGTAGATCAG 2083

Qy     832 LysHisAlaLeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSer 851
Db    2084 CTTGATGCTATCTTGCATGATTCTTTGAAGGCACAGGAGGCTTTGGAATTGATGTCT--- 2140

Qy     852 ArgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGln 871
Db    2141 ------------------------------------------------CCTGGAGAGAAC 2152

Qy     872 GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeu 891
Db    2153 ACTAATATTCTCAAGGCAATGCTAAACTGTGGTTATAAGCCTGATGCTGAGCCCTTTCTT 2212

Qy     892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsn 911
Db    2213 TCAATGATGTTGCAAACCTTCCGCGCATCCAAGTTGCTCGATTTGCGGACTAGATCAAGA 2272

Qy     912 IleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGlu 931
Db    2273 ATATTTATTCCAAATGGAAGAACAATGATGGGATGTTTGGATGAATCCAGAACCTTGGAA 2332

Qy     932 GluAsnGluValHisValGlyPheSer------------SerLysPheArgAspGluGlu 947
Db    2333 TATGGTCAGGTGTTTGTTCAGTTTACTGGTGCTGGACATGGAGAGTTTTCTGACGATTTA 2392

Qy     948 GluSerPhe-----------------------ThrLeuLeuSerAspCysAspValLeu 959
Db    2393 CATCCATTTAATAACAGCAGATCCACCAACAGTAATTTCATTCTGAAGGGAAATGTGGTT 2452

Qy     960 ValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLys 979
Db    2453 GTTGCAAAAAATCCATGCTTGCATCCTGGTGATATTCGTGTTTTAAAGGCTGTAAATGTT 2512

Qy     980 ProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeu 999
Db    2513 CGAGCGCTGCACCACATGGTAGATTGTGTTGTATTCCCTCAGAAAGGAAAAAGACCTCAT 2572

Qy    1000 AlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspPro 1019
Db    2573 CCGAATGAATGTTCTGGGAGTGATTTGGATGGGGATATCTACTTTGTTGCTGGGATCAA 2632

Qy    1020 GluIleValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyr 1039
Db    2633 GACATGATCCCGCCAAGGCAAGTCCAGCCGATGGAATATCCTCCAGCACCCAGCATACAG 2692

Qy    1040 LeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAla 1059
Db    2693 TTGGACCATGAT------------------------------------------------ 2704

Qy    1060 AlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnPro 1079
Db    2705 ------------GTCACAATTGAGGAAGTTGAAGAGTACTTCACCAACTATATTGTGAAT 2752

Qy    1080 AsnPheLeuGlyMetCysThrAsn------------TyrLysGluArgLeuCysTyrIle 1095
Db    2753 GACAGTTTGGGAATCATAGCAAATGCCCATGTCGTATTTGCAGACAGA---------GAA 2803

Qy    1096 AsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuVal 1115
Db    2804 CCTGATATGGCCATGAGTGATCCATGCAAAAAACTTGCTGAGCTCTTTTCAATTGCAGTG 2863

Qy    1116 AspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGlu 1135
```

```
Db       2864 GACTTTCCAAAGACTGGTGTTCCCGCTGAAATACCATCT------CAGTTGCGC------ 2911

Qy       1136 LeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAsp-SerTrpLeuGl 1155
                                            |||                 |||  ||||||:::
Db       2912 ------------------------------CCTAAAGAATACCCAGACTTCATGGATAAG 2941

Qy       1155 yArgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIl 1175
                |||                                                 ||||||||||
Db       2942 CCGG------------------------------------------ACAAGACCAGCTAT 2959

Qy       1175 eAspLysGlu------LeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAspThrGl 1193
              |  ::::::      ||||||:::|||         :::|||
Db       2960 ATCTCAGAAAGAGTTATTGGAAAGCTTTTC-AGGAAAGTGAAG----------------- 3001

Qy       1193 uAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePheLysGluIl 1213
                |||  |||           |||     :::|||      ||||||    ::::::::
Db       3002 -GACAAAGCA------------CCTCAGGCTAGCTCTATCGCGACCTTCACAAGAGATGT 3048

Qy       1213 eSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLysAsnArgIleGlyGl 1233
              ::::  :::|||     :::   :::            ::::::   |||    ||
Db       3049 TGCAAGGAGATCATATGATGCTGATATGGAAGTTGATGGATTTGAAGATTACATTGACGA 3108

Qy       1233 uVal---------GluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSe 1250
              |              :::   ||||||              :::   ::::::   :::
Db       3109 AGCTTTTGACTACAAAACTGAATAT--------------GACAACAAGCTGGGTAATTT 3153

Qy       1250 rLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAlaIleThrPr 1270
                 |||    |||    ::::::                               |||
Db       3154 AATGGACTACTATGGCATAAAAACAGAG-----------------GCTGAAATACTTAG 3195

Qy       1270 oGluAlaMetAspLysSerGlyAlaAsnTyrAsp-----------------------Se 1282
                 :::    |||:::         :::|||                           ||
Db       3196 TGGTGGCATTATGAAGGCATCAAAAACTTTTGACCGCAGAAAAGATGCTGAGGCCATTAG 3255

Qy       1282 rLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrp-- 1301
              |     :::|||   |||                     :::|||        |||
Db       3256 TGTTGCTGTGAGGGCCTTGAGG-----------------AAGGAGGCAAGAGCCTGGTT 3297

Qy       1302 -------------------------AlaLeuLeuArgAlaSerThrAlaPheLysLeuTy 1313
                                       |||    :::||||||        :::    :::
Db       3298 CAAGAGGCGTAATGATATAGATGACATGTTACCAAAGGCTTCGGCTTGGTACCACGTTAC 3357

Qy       1313 rTyrHisLysSer------------------------------------------ProLy 1319
                ||||||    :::
Db       3358 ATATCATCCTACATATTGGGGTTGCTACAATCAGGGGTTGAAAAGAGCTCATTTCATTAG 3417

Qy       1319 sPheValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerAr 1339
                |||  |||    :::       :::   :::         |||    :::     :::||
Db       3418 CTTTCCCTGGTGTGTTTATGACCAGCTAATCCAGATTAAGAAGGACAAAGCACGTAACAG 3477

Qy       1339 gProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAs 1359
              ||||                  :::|||                :::       :::
Db       3478 GCCAGTTCTCAACTTGTCATCTCTCAGGGCTCAACTGAGTCACAGATTAGTG-------- 3529

Qy       1359 pLysLysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAs 1379
                                                 |||:::    |||    |||
Db       3530 ------------------------------TTGAAATGAGATTCCAGTCGAGCGTTAAG 3558

Qy       1379 pProGluValTyrGluValLeuGlyAspAsp 1389
                 :::|||   |||:::|||    |||
Db       3559 CTGATATATATATAATGTAATAGGGTGTGAT 3589
```

RESULT 2
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
;   APPLICANT: Ligon, James M.
;   APPLICANT: Schupp, Thomas
;   APPLICANT: Beck, James J.
;   APPLICANT: Hill, Dwight S.
;   APPLICANT: Neff, Snezanna
;   APPLICANT: Ryals, John A.
;   TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
;   NUMBER OF SEQUENCES: 10
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ciba-Geigy Corporation
;   STREET: 520 White Plains Road, P.O. Box 2005
;   CITY: Tarrytown
;   STATE: NY
;   COUNTRY: USA
;   ZIP: 10591
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/764,233A
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/729,214
;   FILING DATE: 09-OCT-1996
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/258,261
;   FILING DATE: 08-JUN-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Meigs, J. Timothy
;   REGISTRATION NUMBER: 38,241
;   REFERENCE/DOCKET NUMBER: 1506/CIP6
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (919) 541-8587
;   TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 49377 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   ORIGINAL SOURCE:
;   ORGANISM: Sorangium cellulosum
;   IMMEDIATE SOURCE:
;   CLONE: p98/1, pJL3, and pVKM15
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 383..760
;   OTHER INFORMATION: /product= "SorR"
;   OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologou
;   OTHER INFORMATION: the reductase domains of type I PKSs such as eryA from
;   OTHER INFORMATION: Saccharopolyspora erythraea."
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 927..19874
;   OTHER INFORMATION: /product= "SorA"
;   OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKSs tha
;   OTHER INFORMATION: are known to be involved in the synthesis of polyketide
;   OTHER INFORMATION: compounds."
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 942..7115
;   OTHER INFORMATION: /product= "Module 1 of SorA"
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 7203..12884
;   OTHER INFORMATION: /product= "Module 2 of SorA"
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 13455..19616
;   OTHER INFORMATION: /product= "Module 3 of SorA"
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 19871..46318
;   OTHER INFORMATION: /product= "SorB"
;   OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKS gene

```
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19870..24556
;       OTHER INFORMATION:  /product= "Module 1 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  24638..30820
;       OTHER INFORMATION:  /product= "Module 2 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  30881..35446
;       OTHER INFORMATION:  /product= "Module 3 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  35528..40114
;       OTHER INFORMATION:  /product= "Module 4 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  40190..46318
;       OTHER INFORMATION:  /product= "Module 5 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  46851..47891
;       OTHER INFORMATION:  /product= "SorM"
;       OTHER INFORMATION:  /note= "The protein encoded by the sorM gene is highly
;       OTHER INFORMATION:  homologous to the methyltransferase from Streptomyces
;       OTHER INFORMATION:  hygroscopicus that is involved in the synthesis of the
;       OTHER INFORMATION:  polyketide rappamicin."
US-08-764-233A-1

Alignment Scores:
Pred. No.:              8.5e-05        Length:         49377
Score:                  192.00         Matches:        225
Percent Similarity:     31.27%         Conservative:   109
Best Local Similarity:  21.07%         Mismatches:     339
Query Match:            2.60%          Indels:         395
DB:                     1              Gaps:           51

US-09-913-878A-2 (1-1402) x US-08-764-233A-1 (1-49377)

Qy          5 ThrProArgLysArgAsnSerPro-----------------------------ValGlu 14
                :::||||||   :::   ||||||                             :::
Db      10696 TCGCCTCGGTGCAAGCGCTCGCCACGCGCATCACGTCCGCCGAGCAGCTCCGCACCCCAG 10755

Qy         15 GluIleIleAsnArgLeuAsnAsnAspTyrAsnLeuGlyLeuGlnCysValAlaAspThr 34
              |||:::      :::   :::   :::  |||    |||   |||
Db      10756 GAGCTTCCCACCACGATGCCCTCTTCCGCGTCGACTGGAGCGAGCTGCCGAGCC------ 10809

Qy         35 ThrLeuThrProHisArgArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHis 54
                        ||||||||   :::     |||   |||
Db      10810 ------CCACCTCACCGTCTGGAGCCCCAAGCGCCGTCCTTCTCGGCATCGGCGGCCTCG 10863

Qy         55 AspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArgLysAspAspSerLeuAsnGln 74
                          |||   |||:::                     |||
Db      10864 ACCTCGCCCCCGAGGTGCCTCTCGCCCGCGTCGCCGACCTCGCTGCCCTCCAGAGCGCGC 10923

Qy         75 AlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSerSerAsnTrpValProLysAla 94
              :::                  ::::::|||   ::::::|||   :::|||:::
Db      10924 TCGACCAAGGCGCTTCGCCTCCAGGCCTCGTCGTCGTCCCCTTCATGGCTA--------- 10974

Qy         95 HisAlaAspProAspThrLeu------------------------------------- 101
                       :::|||   |||
Db      10975 ------GAACCGCCGACGACCTCATCCAGAGCGCCCACTCCATCACCGCGCGCGCCCTCG 11028

Qy        102 --------------------ProTrpSerLysGluProProArgAlaAlaThrAlaGly 114
                                  |||   :::   |||||||||||:::::: 
Db      11029 CCCTGCTGCAAGCCTGGCTGGCCGACGAACGCCTCGCCTCCTCGCGCCTCGTCCTGCTCA 11088

Qy        115 GlnGlnTrpAlaLeuGlnThrValLeuLeuGluValLeuAsnArgPheMetProProPro 134
                                                            :::|||   |||
Db      11089 CCCGAC-------------------------------GCGCCATCGCTGCCCGCGCCG 11115

Qy        135 AsnAsnThrProGlyArgThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArg 154
                          |||   ||||||            ||||||||   |||          |||
Db      11116 ATGAAGACG---TCAAGGACCTCGCTCACGCCCCTCTCTGGGGGCTCGCACGCTCCGCGC 11172

Qy        155 ProThrSerThrAsnThrLysArgLysAspGluProAlaAsnValThrPheAlaAspPro 174
                 :::|||         |||            :::        |||         |||
Db      11173 AAAGCGAACACCCAGAACTCCCGCTCTTTCTCG---TCGACCTGGACCTCAGTGAGGCCT 11229

Qy        175 Pro-------------------------LysArgSerLeuThrArg---SerAlaThr 184
              |||                          :::   |||:::||||||   |||||||||
Db      11230 CCCAGCACACCCTGCTCGCCGCGCTCGAAACAGGAGAGCGTCACTCGCGTCTCCGCAACG 11289

Qy        185 GlyProPro-----------------IleHis-------------------------- 189
                 |||                   :::|||
Db      11290 GAAAACCCTTCATCCCGAGATTGGCGAATGCACGCTCGAAGGATGAGCTCATCGCCCCGG 11349

Qy        190 -----------GlyAlaAlaIleProLeuLysPhe---ProAspProValAsnThrGly 204
                         ||||||:::|||        |||   |||         :::|||
Db      11350 ACGCGTCCAACTGGCGCCTCCATA---------TTCCGACCAAAGGCAACTTCGACGCGC 11400

Qy        205 SerLys----------------------ArgProSerLeuGluSerGluAsnLeuAsn 216
              |||                         |||||||||   :::::::::   :::
Db      11401 TCACCCTCGTCGACGCCCCTCTAGCCCGTGCGCCCCTCGCACACGGCCAAGTCCGCGTCG 11460

Qy        217 GlnCysThrLysArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaPro 236
                 ||||||
Db      11461 CCGTGCACG--------------------------------------------------- 11469

Qy        237 ProValProIleAlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArg 256
              |||   |||      ||||||:::         |||                  ||||||
Db      11470 CCGCAGCCTTCAATTTCCGCGATGTCCTCGACACCCTTGGTCTGTATCCGGGCGACGCGG 11529

Qy        257 AspProThrAlaThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGln 276
              |||   :::|||            :::|||      :::   |||            :::
Db      11530 GACCGCTCGGCG---GCGAAGGCGCAGGCATCGTTACTGAAGTCGGTCCAGGTGTTTCGC 11586

Qy        277 GlyThrSer--------------TyrGlySerSerValPheSerAlaCysArgHisAsn 291
              |||||||                  :::|||||||||         |||
Db      11587 GGTACACCGTAGGCGACCGGGTGATGGGGATCTTCGGCGCAGCTTGCGGTCCCACGGCCA 11646

Qy        292 GlnSerThrThrGlnSerSerPheGluAlaProProSerGlnProArgGluLysArgPro 311
                    |||         |||   :::|||      |||         |||   |||
Db      11647 TCGCCGACGCCCGCATGATCTGCCCCATCCCCCACGCCTGGTCCTTCGCCCAAGCCGCCA 11706

Qy        312 ValAspAlaThrValPheGluAlaGlyHisLeuIleGluSerProSer------LysGly 329
                                                  |||||||||
Db      11707 GCG--------------------------------TCCCCATCATCTATCTCACCG 11730

Qy        330 ArgThrThrLysSerHisIleAspAsn------GlnProLeuSerSerSerSer------ 345
                 ||||||   |||   |||            |||::::::   ||||||
Db      11731 CCTACTACGGACTCGTCGATCTCGGGCATCTGAAACCCAATCAACGTGTCCTCATCCATG 11790

Qy        346 ---------------GlnGlyGluThrSerPheSerThrTyrTyrGluSerPheProSer 360
                             |||         ||||||:::::         |||   |||
Db      11791 CGGCCGCCGGCGGCGTCGGGACGGCCGCCGTTCAGCTCGCACGCCACCTCGGCGCCGAGG 11850

Qy        361 SerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLeu------AlaArg 378
              |||                  :::|||   |||::::::   :::|||       ||||||
Db      11851 TCT---------------TTGCCACCGCCAGCGCAGGGAAGTGGAGCGCTCTCCGCGCGC 11895

Qy        379 SerGluGluSerAlaArgSerGlnValGlnValHisAla-ProValValAlaAlaArgLe 398
              |||      :::   |||:::      :::   |||      |||   :::   |||   ||
Db      11896 TCGGCTTCGACGACGCGCACCTCG---CGTCCTCACGTGACCTGGACTTCGAGCAGCACT 11952

Qy        398 uArgAsnIleTrpProLysPheProLysTrpLeuHisGlu------------------- 411
              |          |||         |||
Db      11953 T------------CCTGCGCTCCACGCATGGGCGCGGCGTGGATGTCGTCCTCGACTGCT 12000

Qy        412 ---------------------------AlaProLeuAlaValAlaTrpGluValThrAr 422
                                         ||||||   |||         |||   :::   ||
Db      12001 TGGCACGCGAGTTCGTCGACGCTTCGCTGCGCCTCATGCCGAGCGGTGGACGCTTCGTCG 12060

Qy        422 g-LeuPheMetHisCysLysValAspLeu---GluAspGluSerLeuGlyLeuLysTyrA 441
              | :::            |||   |||:::   |||      :::::::::|||:::   |||
```

```
Db      12061 AGATG---------GGCAAGACGGACATCCGTGAGCCCGACGCGGTCGGCGTCGCCTACC 12111

Qy        441 spProSerTrpSerThrAlaArgAspValThrAspIle--------------------- 453
                               |||   |||:::   :::
Db      12112 CTGGTGTCGTTTACCGCGCCTTCGACCTCATAGAGGCCGGACCGGATCGAATAGAGCAGA 12171

Qy        453 ------------------------------------------------------------ 453

Db      12172 TGCTCGCAGAGCTGCTCAGCCTCTTCGAGCGCGGTGCGCTTCGTCCGCCGCCCATCACAT 12231

Qy        454 --TrpLysThrLeuTyrArgLeuAspAlaPheArg-----------------------G 465
                |||          :::          ||||||||                       |
Db      12232 CTTGGGACATCCGTCATGCCCCCCAGGCCTTTCGCGCGCTCGCTCAGGCGCGGCATGTTG 12291

Qy        465 lyLys---------------ProPheProGluLysProProAsnAspValPheValThrA 480
              |||||               |||   |||   |||         |||   :::|||
Db      12292 GGAAGTTCGTCCTCACCATACCCCGTCCCATAGACCCCGAAGGCACCGTCCTCATCACGG 12351

Qy        480 laMetThrGly------------------------------------------------- 483
                 |||||
Db      12352 GAGGCACCGGCACGCTAGGAGCCCTGGTCGCACGCCATCTCGTCGCAAGACACGGCGCCA 12411

Qy        483 ------------------------------------------------------------ 483

Db      12412 AGCACCTGCTTCTCACGTCGAGGCAGGGCGCGCACGCTCCGGGCGCCGAGGCCTCGCGAA 12471

Qy        484 --AsnPheGluSerLysGlySerAlaValValLeuSerAlaValLeuAspTyrAsnPro- 502
                      |||:::   |||::::::|||   |||   |||         :::|||
Db      12472 CCGAGCTCGAAGCGCTGGGGGCCTCTGTCACACTTCGCGCGTGCGACGCGGCCGACCCAC 12531

Qy        503 --------------------AspAsnSerProThrAla---ProLeuTyrLeuValLysL 515
                                  |||:::   |||:::|||   ||||||      |||
Db      12532 GCGCCCTCCAAGCCCTCTTGGACTCCATCCCGAGCGCTCACCCGCTCACCGCCGTCGTCC 12591

Qy        515 euLysProLeuMetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgP 535
                                 :::   |||      |||               |||:::|||
Db      12592 ACGCCGCGGGCGCCCTCGACGACGGC------CTGCTCGGCGCCATGAGCCCCGAGCGCA 12645

Qy        535 hePheGluIle-LeuIlePro--------------------------SerProThrSer 545
                 :::   |||   |||                           |||
Db      12646 TCGACCGCGTCTTTGCCCCCAAGCTCGATGCTGCTTGGCACTTGCATGAGCTCACCCAAG 12705

Qy        546 ThrSerProSerValProProValValSerLys-------------------------- 556
              |||||||||||      |||         |||:::
Db      12706 ACAAGCCCCTCGCCGCCTTCGTCCTCTTCTCGTCCGCTGCTGGCGTCCTTGGTAGTCCAG 12765

Qy        557 ------------------------GlnProGlyAlaValGluGluValIleGlnTrpLeu 568
                                         |||   ::::::         |||      :::
Db      12766 GTCAGTCGAACTACGCCGCGGCCAATGCCTTCCTCGATGCGCTCGCGCATCACCGGCGTG 12825

Qy        569 ThrMetGly---------------GlnHis-SerLeuValGlyArg------------- 578
                 |||                  |||    |||:::||||||
Db      12826 CCCACGGGCTCCCGGCCTCCTCGCTCGCATGGGGCTATTGGGCCGAGCGCAGTCGAATGA 12885

Qy        579 ----------GlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLys-------- 592
                        |||   |||      |||:::   :::|||      |||
Db      12886 CCGAGCACCTCAGCGCCGCCGATGTTTCTCGCATGAGGCGCGCCGGCGTCCGGCCCCTCG 12945

Qy        593 ------------------ProLeuArgGluPheGlnLeuArgAlaGlu---------As 603
                                ||||||||         |||   |||
Db      12946 CCACAGACGAGGCGCTCTCCCTCTTCGATGCGGCTCTCTTGCGGCCCGAGCCCGCCCTGG 13005

Qy        603 pProLysProIleIleLysGluArgValHisPhePheAlaGluThrGlyIleThrPheAr 623
               |||:::   :::   :::||||||                                  ||
Db      13006 TCCCCGCACGCTTCGACGTGAACGCGCT------------------------------CG 13035

Qy        623 gProAspValPheLysThrArgSerValValProAlaGluGluProValGluGlnArgTh 643
              |   :::         :::            |||||||||      |||      :::|||||
Db      13036 GCGCGAATGCCGACGAGGTGCCCCCGCTGTTCCAGCGTCTCGTCCGCGCTCGCGTCGCAC 13095

Qy        643 rGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAspAsnAsnThrTrpGl 663
              |:::                                                  :::::
Db      13096 GCAA----------------------------------------GGCCGCCAGCAATACCG 13116

Qy        663 nProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTyrAlaIleMe 683
              :|||   |||                 |||         |||      ::::::         ::
Db      13117 CCCTGGCCTCTTCACTCTCTCAGCGCCTCTCCTCCCTCCCGCCCGCAGAAAGCGAGCGCT 13176

Qy        683 t------ThrLeuGluProHisGlnIleArgHisHisLysThrAspLeuLeu-------- 698
              :                |||||||:::   ||||||               |||
Db      13177 TCCTCCTCGATCTCGTCCGCACCGAAGCCGCCACCGTCCTCGGCCTCGCCTCATTCGAAT 13236

Qy        699 -------SerProSerGlyThr-GlyGluValMetAsnAspGlyVal------------- 711
                     |||||||||     :::      |||:::      |||   :::
Db      13237 CGCTCGATCCCCATCGCCCTCTCCAAGAGCTCGGCCTCGATTCTCTTATCGCTCTCGAGC 13296

Qy        712 -----GlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyA 730
                   ||||||::::::   :::            |||::::::      |||   |||   |
Db      13297 TCCGAGGTCGACTCGCCGCGGCCACCGGGCTGCGACTCCAACCTACTCTC---CTCTTCG 13353

Qy        730 spValProSer---AlaValGlnGlyArgPhe--------------------------- 739
               ||   |||:::      |||         ||||||
Db      13354 ACTATCCAACCCCGGCTGCACTCTCACGCTTTTTCACGACGCAGTTCTTCGGGGAAACCA 13413

Qy        740 -------------------------------GlySer---------------------- 741
                                             ||||||
Db      13414 CCGACCGTCCCGCAGCGCCGCTCACCCCGGCGGGAAGCGAAGACCCTATCGCCATCGTGT 13473

Qy        742 ----------------------------------AlaLysGlyMetTrpValIleAspV 750
                                                :::   :::|||      :::     :
Db      13474 CGATGAGCTGCCGCTTCCCTGGCGACGTGCGCACGCCCGAGGATCTCTGGAAGCTCTTGC 13533

Qy        750 alAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGlnArgLysTrpGluC 770
              ::|||              :::|||   |||      ::::::|||      |||   |||
Db      13534 TCGAT---------GGGAAAGATGCCATCTCCAGCTTTCCCCAGAATCGCGGTTGGAGTC 13584

Qy        770 ysAspPheValAsp 774
                |||   :::|||
Db      13585 TCGATGCGCTCGAC 13598
```

RESULT 3
US-09-252-991A-10214
; Sequence 10214, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10214
;   LENGTH: 4026
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10214

```
Alignment Scores:
Pred. No.:               9.92e-06      Length:          4026
Score:                   182.00        Matches:         273
Percent Similarity:      33.31%        Conservative:    192
Best Local Similarity:   19.56%        Mismatches:      473
Query Match:             2.46%         Indels:          465
DB:                      4             Gaps:            65

US-09-913-878A-2 (1-1402) x US-09-252-991A-10214 (1-4026)

Qy         86 AlaSerSerAsnTrp-ValProLysAlaHisAlaAspProAspThrLeuProTrpSerLy 105
              ::::::|||   |||   ||||||:::|||            :::|||   |||       |||   ::
Db        401 TCGAACAGCCGCTGGAGTTCCTCAAGCGGCAGAAGAACCTGC-ACCGGGTCCTGGCCGAG 459
```

```
Qy    105 sGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeuGl 125
Db    460 GGACAGTTCAAGGGCGAAGGCCTGTCGAACACCCTGGCCTATCTCT-------------- 505

Qy    125 uValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArgTh 145
Db    506 -------------------------------------------GGGACAACTGGTATCGCAAG 525

Qy    145 rLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAspGl 165
Db    526 TCGCGCCGGGTGATGCAGCGCGCCTTCTCCTACGAGTCGCGGCAACAGGTCACCGAGGCG 585

Qy    165 uProAlaAsnValThrPheAlaAspPro---ProLysArg-SerLeuThrArgSerAlaT 184
Db    586 GTGCCCGAGCTGAAGATGGGCAACGCCATGACCGCGCCGACGCCCTACGACGCGCAGCCG 645

Qy    184 hr------------GlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspP 200
Db    646 CAATACGCGCCGATTCCCGGCCTGCCGCTGGTCAGCCAGTTCCGCTA---------TCCG 696

Qy    200 roValAsnThrGlySerLysArgProSerLeuGluSerGluAsn---------LeuAsnG 217
Db    697 CTGTGGACGCCAAGCCCATCGAGCCGCCGCAGGACGTCAAGCTGGCCGGTTCCTCCAGC 756

Qy    217 lnCysThrLysArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProP 237
Db    757 GAGTTCATCAACGTCGTGCCGGGCAATGTCTACATCCCGCTGGGCAAGCTGAAGCCGGGC 816

Qy    237 roValProIleAlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgA 257
Db    817 CTGTACCTGGTCGAGGCGCTTGTCGGCAAGTACCG---CGCCACCACCGTGGTGTTCGTC 873

Qy    257 spProThrAlaThrGlyHisArg---ArgAlaAspGlnValAspSerPheAspThrSerG 276
Db    874 TCCAACAGCGTGGCGGTGAGCAAGGTCGCCGGCGACGAGCTGCTGGTCTGGACCGCACGC 933

Qy    276 lnGlyThrSerTyrGlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrG 296
Db    934 AA------------------------------GCACGAAGGCACCCCGGTGCCCGACACC 963

Qy    296 lnSerSerPheGluAlaProProSer---------------GlnProArgGluLysArgP 311
Db    964 AAGGTGCTGTGGACCGACGGCCTGGGCGTGATGAGCAGCGGCAATACCGACGCCGACGGC 1023

Qy    311 roValAspAlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgT 331
Db   1024 CTGCTGCGGCTCAAGCACGCCAGTCCGGAGCGCTCCTACGT------GATCGGCG-AGGA 1076

Qy    331 hrThrLysSerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerP 351
Db   1077 CCGCGAGG-----------------------GCGGGGTCTTCGTCTCCGAGAACTTCTA 1112

Qy    351 heSerThrTyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluPro- 370
Db   1113 TTACGACAGCGAGATCTACGACACCAAGATCT---------------ACGCCTTCACCGA 1157

Qy    371 --SerArgSerAsnGlyLeuAlaArg------------------SerGluGluSerAlaA 384
Db   1158 CCGCCCGCTGTACCGGCCGGGCGACTGGGTCTCGTTGAAGATGGTCGGCCGCGAGTTCAA 1217

Qy    384 rgSerGlnValGlnValHisAlaPro-----------------------------ValV 394
Db   1218 GGACGCCCGGCAATCCCAGGCCGCCGCCAGCGCGCCGGTGCGCCTGAGCGTGATCGACGC 1277

Qy    394 alAlaAlaArgLeuArgAsnIleTrpProLysPheProLysTrpLeuHisGluAlaProL 414
Db   1278 CAGCGGCACGGTATTGCAAAGCCTGGACCTCC---GCTTCGACGCGAAGAGCGGCGCCAA 1334

Qy    414 euAlaValAlaTrpGluValThrArgLeuPheMetHisCysLysValAspLeuGluAspG 434
Db   1335 TGGCCGCTTCC-----------------------AACTGCCGGAG-AACGCCGTCGCCG 1369

Qy    434 luSerLeuGlyLeuLysTyrAsp---------ProSerTrpSerThrAlaArgAspValT 451

Db   1370 GCGGCTACGAGTTGCGCTTCGACTACCGCGGGCAGACCTACAGCAGCGCCTTCCGCGTCG 1429

Qy    451 hrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArgGlyLysPro---------- 467
Db   1430 CCGAATACATCAAGCCGCACTTCGAAGTCGCCCTCGACCTGGCCAAGCCGGACTTCAAGA 1489

Qy    468 ----------------------------------PheProGluLysProPro------- 473
Db   1490 CCGCCGAGCCGGTGAAGGGCGAGATCGTCCTGCTCTATCCGGACGGCAAGCCGGTGGCCA 1549

Qy    474 ------------------------------------------------AsnAspValP 477
Db   1550 ACGCGCGCCTGCAACTGAGCCTGCGCGCCCAGCAGTTGTCGATGGTGGACAACGAACTG- 1608

Qy    477 heValThrAlaMetThrGlyAsnPheGluSerLysGlySerAlaValValLeuSer---- 495
Db   1609 -------CAGTATCTCGGCCAGTTCCCGGTGGAACTGAGCAGCACCGAACTGACCACAG 1660

Qy    496 --------------AlaValLeuAspTyrAsnProAspAsnSerProThr---------- 507
Db   1661 ACGGCAAGGGCCGCGCCGCCATCGAATTGCCGCCGGCCGAGAAACCCAGTCGCTACATGC 1720

Qy    508 -------------------AlaProLeuTyrLeuValLysLeu---LysProLeuMetP 520
Db   1721 TGACCATCTTCGCCAGCGACGGCGCCGCCTACCGGGTCAAGACCAGCAAGGAGATCCTCA 1780

Qy    520 heGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhe------------- 535
Db   1781 TCGAGCGCGGCGCCGCGCGCTACCGCCTGAGCGCGCCGCAGCGCTTCAGCGCCGCCGGCG 1840

Qy    536 -----------PheGluIleLeuIleProSerProThrSerThrSerProSerValProP 552
Db   1841 AGAAGGTCGAGTTCAGCTATGCCAGCGAACAGCCGACCCCGCTCAAGCCGAGC------- 1893

Qy    552 roValValSerLysGlnProGlyAlaValGluGluValIleGlnTrpLeuThrMetGlyG 572
Db   1894 ----------------------------------AGCTACCAGTGGATCCGCCTGGAGG 1918

Qy    572 lnHisSer---------------LeuValGlyArgGlnTrpArgAlaPhePheAlaLysA 587
Db   1919 ACCGCGCCACCGACAGCGGTCCGGTCGCCGACGGCCGCTTCGCGCTGACCTTCGAGCGCC 1978

Qy    587 spAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysProI 607
Db   1979 CCGGTACCTATTCCGTGGAGCTGCGC------------------GACGACAAGGGCCAGC 2020

Qy    607 leIleLysGluArgValHisPhePheAlaGluThrGlyIleThrPheArgProAsp---- 625
Db   2021 TCCTTGGCGCCACCGGCCACTCGGTGAGCGGCGAGGGCGTGAAGTCGGTGCCGGGCACCG 2080

Qy    626 --------ValPhe---------------------------LysThrArgSerValValP 634
Db   2081 TCGAGGTGGTGTTCGACAAGCCCGAGTACCGCACCGGCGAGGAAGCCTCCGCACTGATCA 2140

Qy    634 roAlaGluGluProValGlu------------------GlnArgThrGluPheLysValS 648
Db   2141 CCTTCCCCGAGCCGGTGGAGGATGCGCTGTTGTCCCTGGAGCGCGACAAGGTCGAGGCCA 2200

Qy    648 erGlnMetLeu------------AspTrpLeu---LeuGlnLeuAspAsnAsnThr---- 661
Db   2201 CCGCCCTGCTGTCCAAGGGCGCCGACTGGCTGCGCCTGGAAAAACTCAATCCGACCCAGT 2260

Qy    662 --------TrpGlnProHisLeuLysLeuPheSer------------------------- 670
Db   2261 ACCGCGTGTGGATCCCGGTGCGCGAGGAGTTCTCACCGAACCTGACCTTCTCGGTGCTCT 2320

Qy    671 -----------------------------------ArgIleGlnLeuGlyLeuSerLysT 679
Db   2321 ACACCAAGGGCGGCGACTACAGCTTCCAGAACGCCGGGATCAAGGTCGGCATGCCCCAGG 2380

Qy    679 hrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThr---------- 695
```

```
Db    2381 TAGAGATCGACATCGCCACCGACAAGGAGCGCTACGAACCGGGCGAGACGGTGACCGTGA 2440

Qy     696 -----------------------AspLeuLeuSerProSerGlyThrGlyGluValMetA 708
                                        :::|||         |||      |||::::
Db    2441 CCCTGGCCACGCGCTTCGCCGGCAAACCGGTTTCCAGCCACCTGACCGTCAGCGTGGTGG 2500

Qy     708 snAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArgAspValLeuGlyL 728
           ::::   |||     :::           :::|||   |||   |||   ||||||
Db    2501 ATGAGATGGTCTATGCCCTGCAGGCGGAGATCGCCCC-CGGCATCGACCAGTTCTT---- 2555

Qy     728 euGlyAspValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValI 748
                   :::|||||||||   |||   |||         |||
Db    2556 --------CTACCATCCGCGGCGCAACAACGTGC-GCACCAGCG---------------- 2590

Qy     748 leAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSerGln---ArgL 767
                                                    :::||||||
Db    2591 ----------------------------------CCAGCCTGGCCTTCATCAGCTACGA 2615

Qy     767 ysTrpGluCys--------------------AspPheValAspLysHisGlnArgThrL 780
             |||   |||                     |||      ::::::   :::|||   :
Db    2616 CGTGGCGTTGCCCGGCAGCACCAGCGCACCCGGACGG-GCGAACCGCAGCGAGCGCGGCG 2674

Qy     780 euGluVal------ArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnL 798
           :::::|||       |||            ::::::   :::|||            :::|
Db    2675 TCAAGGTACTGGAGCGGCCGCGCCGCGAGGATGTCGACACCGCCGCCTGGCAGCCGGAGC 2734

Qy     798 euLeuProValLeuGluAspArgAlaArgAspLysValLysMet---------------A 813
           ||:::         :::   :::|||            :::|||               |
Db    2735 TGGTCACCGACGCCCAGGGCAAGGCCAGCTTCAGCTTCCGCATGCCCGACTCGCTGACCC 2794

Qy     813 rgGlnAlaIleGlyAspArgLeuIleAsnAsp---------------LeuGlnArgGlnP 828
           ||      |||      |||   |||:::|||                   ::::::||||
Db    2795 GCTGGCGGATCACCGCGCGGGCCATCGACGACAACGGCCAGGTCGGCCAGAAGAAGCAGT 2854

Qy     828 heSerGluGlnLysHisAlaLeu------------AsnArgProValGluPheArgGln- 843
           ||         :::      |||            :::   |||       ||||||||
Db    2855 TCCTTCGCTCGGAGAAGCCGCTGTACCTGAAGTGGAGCGGGCCGACTCGCTTCCGCCAGG 2914

Qy     844 -----------------------TrpValTyrGluSerTyrSerSerArgAlaThrArgV 856
                                  :::|||:::
Db    2915 GTGACCAGCCCGACCTCGGCCTGTTCGTCTTC---------------------------- 2946

Qy     856 alSerHisGlyArgValPro------------PheLeuAlaGlyLeuProAspSerGln- 871
             :::   |||      |||               |||:::|||    |||   ||||||
Db    2947 --AACCAGGGCGAACAGCCGGTCAAGGCCGAGCTGCTCAGCGGGCCGCCTGGCAGCCAGC 3004

Qy     872 --GluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrL 891
                :::||||||                       :::   |||      |||::: |||
Db    3005 GCAGCCAGACCCTGGAGCTGGCCAAGGGAGTCAACTACATTCCGCTGGCGCAACAGCCGC 3064

Qy     891 euGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuA 911
           ||   |||      |||      :::   |||:::         ::::::   :::|||
Db    3065 TCAGCGATGGCGACTGGAGCGCCGAGCTGCGCCAGGACGGGCAGGTCCGCGACCGCCTGG 3124

Qy     911 snIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuG 931
             :::|||                        ::::::|||||     |||   |||  :
Db    3125 CGGTGCGCTTCAAC------------------CTGCTGGCCGACGGCTGGCAGGTGGAGC 3166

Qy     931 luGluAsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGluSerPheT 951
           ::
Db    3167 AG---------------------------------------------------------- 3168

Qy     951 hrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspI 971
                :::   :::      :::      |||   :::|||          |||:::|||
Db    3169 -----GTGCAGAACCTCAGCCTGGCGGCGGCGAGCAACCCGCTGCAACTGCCGGCGGACG 3223

Qy     971 leGlnArgValArg-----------------------AlaValPheLysProGluLeuH 983
             :::   ||||||                        |||   :::         |||
Db    3224 CGCGCGACGTACGCCTGCGCCTGGCCGATGGCCCGGCCGCGGCCTAC---------CTGG 3274

Qy     983 isSerLeuLysAspValIleIlePheSerThrLysGlyAsp------------------- 996
             :::|||   |||::::::   :::         |||
Db    3275 GCAATCTCGATGATCTGCTGGAGTATCCCTACGGCGGCGTCGAGCAGACCGCCAGCCAAC 3334

Qy     997 --ValProLeuAla---------LysLysLeuSerGlyGlyAsp---------------- 1007
             :::||||||:::               |||:::||||||:::
Db    3335 TGCTGCCGCTGAGCATCGCCTATCCGGCGCTGGCCGGCGGCGAGCCGCGGATTCGCGACC 3394

Qy    1008 ---------------------------------------------TyrAspGlyAspMetA 1013
                                                         |||       |
Db    3395 GCCTGCGCCTGATCATGCAGAACAGTCGCCTGCGCCTGGTGCAGATGGCCGGCCCGGACG 3454

Qy    1013 laTrpValCysTrpAspProGluIleValAspGly------PheValAsnAla------- 1028
           |||||      |||       |||   ||||||||        |||:::   |||
Db    3455 CCTGGTTCGCCTGGTGGGGTGGCGACGTCGATGGCGACGCTTTCCTCACCGCCTACGCCT 3514

Qy    1029 ---------------------------------------GluMetProLeuGluProAspL 1036
                                                 ::::::|||    |||
Db    3515 ACTACGCCGACTGGTACGCCAGCCGCGCACTGGAGATCCAACTGCCGGCCGAA------- 3567

Qy    1036 euSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyT 1056
                                           ::::::::::::::        :::
Db    3568 ------------------------------CACTGGCAGCGCATCCTCGAGCCCTAC---- 3594

Qy    1056 hrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheA 1076
                      |||||:::    |||          :::|||::::::            |
Db    3595 ----------GCCAAGCAGGCGACGCAGACCCCGCTGCTGCAGCGTGCGCTGATCCTGG 3643

Qy    1076 laLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleA 1096
           ||
Db    3644 CTTTC------------------------------------------------------- 3648

Qy    1096 snAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValA 1116
                      :::            :::   ::::::::|||:::|||     |||:::
Db    3649 -----------GCCCGCGACATGCAGTTGCCGGTGAACACCCTGCTCGGCGGCCTGCTCA 3697

Qy    1116 spGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluL 1136
             :::|||   |||                                     ||||||:::
Db    3698 A-CGATCTGGCCAA--------------------------------------CGCCGGCAAA 3720

Qy    1136 euLeuGlyGlyAlaLeuSerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyA 1156
                |||   |||   :::         |||:::    ::::::|||      |||
Db    3721 GGGCAGGGCGCGCGCGGACC---------CCGCTG---GAGGCCGACGACGGCTGGTGC 3768

Qy    1156 rgGlyGluProThrHisIleIleAspTyrLeuLysPheSerIleAlaArgProAlaIleA 1176
             |||:::|||         :::                   :::            :::
Db    3769 TGGGCGACCCGGACTCGGCCGTCGGCCTGGCCGCGGCGCGAGTGCTGGCGGTGGACCTTG 3828

Qy    1176 spLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLysAsp------------ 1191
             ::::::|||               |||       ||||||:::
Db    3829 CACGGCAACTGCGGGTCGCGGTGCCCGCGCCGCTGGCCGCCCAGGCCGAGACGGCGACGC 3888

Qy    1192 --------ThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrP 1209
                      |||   |||      |||   |||       ||||||||
Db    3889 AGCGCCTGCGTGAGGCCGGCTTGCCGTTCACCGATGCCTTGCTGGCCTCT---------- 3938

Qy    1209 hePheLysGluIleSerAspLysSerArgSerSerAlaLeuLeu 1223
                :::         |||      ::::::||||||||||||
Db    3939 -----CGCTCGGCCGTGGATGGCCAGCAGGCCAGCGCTCTGCTG 3977

RESULT 4
US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
```

```
;    NUMBER OF SEQUENCES: 22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Ciba-Geigy Corporation
;      STREET: 7 Skyline Drive
;      CITY: Hawthorne
;      STATE: NY
;      COUNTRY: USA
;      ZIP: 10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/258,261B
;      FILING DATE: 08-JUN-1994
;      CLASSIFICATION: 800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/457,205
;      FILING DATE: 01-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Elmer, James Scott
;      REGISTRATION NUMBER: 36,129
;      REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 919-541-8614
;      TELEFAX: 919-541-8689
;  INFORMATION FOR SEQ ID NO: 6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 28958 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;    MOLECULE TYPE: DNA (genomic)
;    HYPOTHETICAL: NO
;    ANTI-SENSE: NO
US-08-258-261B-6

Alignment Scores:
Pred. No.:              0.000564    Length:         28958
Score:                  178.50      Matches:        213
Percent Similarity:     29.87%      Conservative:   89
Best Local Similarity:  21.07%      Mismatches:     399
Query Match:            2.41%       Indels:         312
DB:                     1           Gaps:           42

US-09-913-878A-2 (1-1402) x US-08-258-261B-6 (1-28958)

Qy      38 ProHisArgArgLys-----------------------------GluLeuAlaGluSer 47
Db    5663 CCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCC---TCG 5719

Qy      48 AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg 67
Db    5720 CCGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGA 5779

Qy      68 LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSer 87
Db    5780 TCGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCT 5839

Qy      88 Ser---------AsnTrpVal--------------------------ProLysAla--- 94
Db    5840 TCCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCA 5899

Qy      95 ------------HisAlaAspProAspThr------------------LeuProTrpSer 104
Db    5900 CCTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACC 5959

Qy     105 LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeu 124
Db    5960 ACGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCG 6019

Qy     125 GluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArg 144
Db    6020 ACGGCGTACGTGCCCTTC---ACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTC 6076

Qy     145 ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp 164
Db    6077 ACGCTGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCAT 6136

Qy     165 GluProAla---------AsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
Db    6137 GCGCCTTCCTTCCCACCCTCCGCAAGGACGCGACGACGCCGAGGCGTTCACCGCCGCGC 6196

Qy     182 SerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProVal 201
Db    6197 TCGGCGCTCTCCACTCCG-----------------------CAGGCATCACACCCGACT 6232

Qy     202 AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
Db    6233 GGAGCGCTTTCTTCGCCCCCTTCGCTCCAC------------------------------ 6262

Qy     222 AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValPro---Ile 240
Db    6263 ---------------------------GCAAGGTCTCCCTCCCCACCTATGCCTTCCAGC 6295

Qy     241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
Db    6296 GCGAGCGCTTCTGGCCCGACGCCTCCAAGGCACCCGGCG------------CCGACGTCA 6343

Qy     261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
Db    6344 GCCACCTTGCTCCGCTCGAGGGGGGGCTCTGGCAAGCCATCGAGCGCGGGGACCTCGATG 6403

Qy     281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
Db    6404 CGCTCAGCGGTCAGCTCCACGTGGACGGCGACGAGCGGCGCGCCGCGCTCGCCCTGCTCC 6463

Qy     301 AlaProProSer--------------------GlnProArgGluLysArgProValAsp 313
Db    6464 TTCCCACCCTCTCGAGCTTTCGCCACGAGCGGCAAGAGCAGAGCACGGTCGACGCCTGGC 6523

Qy     314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys 333
Db    6524 GCTACCGTA-----------------------TCACCTGGAAGCCTCTGACCACCGCCG 6559

Qy     334 SerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThr 353
Db    6560 AAACACCCGCCGACCTCGCCGGCACCTGGCTCGTCGTCGTGCCGGCCGCTCTGGACGACG 6619

Qy     354 TyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSer 373
Db    6620 ---ACGCGCTCCCCTCCGCGCTCA-----------------------CCGAGGCGCTCA 6652

Qy     374 AsnGlyLeuAlaArgSerGluGlu--------------------------SerAlaArg 384
Db    6653 CCCGGCGCGGCGCGCGCGTCCTCGCCTTGCGCCTGAGCCAGGCCCACCTGGACCGCGAGG 6712

Qy     385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys 404
Db    6713 CTC--------------TCGCCGAGCATCTGCGCCAGGCTTGCGCCGAGACCGCCCCGA 6757

Qy     405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424
Db    6758 TTCGCGGCGTGCTCTCGCTCCTCGCCCTCGACGAGCGCC------CCCTCGCAGACCGTC 6811

Qy     425 MetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrp 444
Db    6812 CTGCCCTGCCCGCCGGACTCGCCCTCTCGCTTTCTCTCGCTCAAGCCCTCGGCGACCTCG 6871

Qy     445 SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg 464
Db    6872 ACCTCGAGGCGCCCTTGTGGTTCTTCACGCGCGGCGCCGTCTCCATTGGACACTCTGACC 6931

Qy     465 GlyLysProPheProGluLysPro------------------------------------ 472
```

```
Db   6932 CCCTCGCCCATCCCGCCCAGGCCATGACCTGGGGCTTGGGCCGCGTCATCGGCCTCGAGC 6991
Qy    473 ---------ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer------ 487
Db   6992 ACCCCGACCGGTGGGGAGGTCTCGTCGACGTCTGCGCTGGGGTCGACGAGAGCGCCGTGG 7051
Qy    488 -----------------LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
Db   7052 GCCGCTTGCTGCCGGCCCTCGCCGAGCGCCACGACGAAGACCAGCTCGCTCTCCGCCCGG 7111
Qy    502 ProAspAsn---------SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu 518
Db   7112 CCGGACTCTACGCTCGCCGCATCGTCCGCGCCCCGCTCGGCGATGCGCCTCCCG------ 7165
Qy    519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIle 538
Db   7166 -----------------CGCGCGACTTCACGC-------------------------- 7180
Qy    539 LeuIleProSerProThrSerThrSerProSerValPro---ProValValSerLysGln 557
Db   7181 ------CCGGAGGCACCATTCTCATCACCGGCGGCACCGGCGCCATTGGCGCTCACGTCG 7234
Qy    558 Pro---GlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
Db   7235 CCCGATGGCTCGCTCGAAGAGGCGCTCAGCACCTCGTCCTCATCA-------------- 7279
Qy    577 GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596
Db   7279 ------------------------------------------------------------ 7279
Qy    597 PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla 616
Db   7280 ---------GCCGCCGAGGCGCCGAGGCCCCTG-------------------------- 7303
Qy    617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
Db   7304 -----------------GCGCCTCGGAGCTCCACGACGAGCTCTCGGCCCTCGGCGCGC 7345
Qy    637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln 656
Db   7346 GCACCACCCTCGCCGCGTGCGATGTCGCCGACCGGAATGCTGTCGCCACGCTTCTTGAGC 7405
Qy    657 LeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle------GlnLeu 674
Db   7406 AGCTCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGAGCGGCATCGAACACC 7465
Qy    675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis------------------ 688
Db   7466 ACGCTCCGCTCGACGCCACCTCTTTCAGGGATCTCGCCGAGGTTGTCTCCGGCAAGGTCG 7525
Qy    689 GlnIleArgHisHisLysThrAspLeuLeuSer------ProSerGlyThrGlyGluVal 706
Db   7526 AAGGTGCAAAG-CACCTCCACGACCTGCTCGGCTCTCGACCCCTCGACGCCTTTGTTCTC 7584
Qy    707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg--------- 723
Db   7585 TTTTCGTCCGGCGC-GGCCGTCTGGGGCGGCGGACAGCAAGGCGGCTACGCGGCCGCAAA 7643
Qy    724 ---------------------------------------------------AspValLeu 726
Db   7644 CGCCTTCCTCGACGCCCTTGCCGAGCATCGGCGCAGCGCTGGATTGACAGCGACGTCGGT 7703
Qy    727 GlyLeuGlyAspVal------------------------------------------Pro 732
Db   7704 GGCCTGGGGCGCGTGGGGCGGCGGCGGCATGGCCACCGATCAGGCGGCAGCCCACCTCCA 7763
Qy    733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet------------TrpValIl 748
Db   7764 ACAGCGCGGTCTGTCGCGGATGGCCCCCTCGCTTGCCCTGGCGGCGCTCGCGCTGGCTCT 7823
Qy    748 eAspValAspAspThr------------GlyAspGluAspTrpIleGluThrTyrProSe 764
Db   7824 GGAGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTC 7883
Qy    764 r---------------GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
Db   7884 GTTCAGCGCCGCTCGCCCCCGCCCGCTCCTGCGCGAT---TTGCCCGAGGCGCAGCGCGC 7940
Qy    779 rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLe 799
Db   7941 TCTCGAGACCAGCGAAGGCGCGTCCTCCGAGCATGGCCCGGCC----------------- 7983
Qy    799 uProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspAr 819
Db   7984 -CCCGACCTCCTCGACAAGCTCCGGAGCCGCTCGGAGAGCGAGCAGCTT---------CG 8033
Qy    819 gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
Db   8034 TCTGCTCGTCTCGCTGGTGCGCCACGAGACGGCCCTCGTCCTCGGCCACGAAGGCGCCTC 8093
Qy    839 ----------------------------------ValGluPheArgGlnTrp------- 844
Db   8094 CCATGTCGACCCCGACAAGGGCTTCCTCGATCTCGGTCTCGATTCGCTCATGGCCGTCGA 8153
Qy    844 ------------------------------------------------------------ 844
Db   8154 GCTTCGCCGGCGCTTGCAACAGGCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGA 8213
Qy    845 -----------ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV 861
Db   8214 CCATCCCTCTCCTCATCGAGTCGCGCTCTTCTTGCGCGACTCGC---------TCGCCCA 8264
Qy    861 alProPheLeuAlaGlyLeuProAspSer 870
Db   8265 CGCCCTCGGCACGAGGCTCTCCGTCGAGC 8293
```

RESULT 5

```
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/456,837
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
```

```
;       REGISTRATION NUMBER:  36,129
;       REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8614
;       TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  28958 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-456-837-6

Alignment Scores:
Pred. No.:              0.000564      Length:        28958
Score:                  178.50        Matches:       213
Percent Similarity:     29.87%        Conservative:  89
Best Local Similarity:  21.07%        Mismatches:    399
Query Match:            2.41%         Indels:        312
DB:                     1             Gaps:          42

US-09-913-878A-2 (1-1402) x US-08-456-837-6 (1-28958)

Qy       38 ProHisArgArgLys-----------------------------GluLeuAlaGluSer 47
             |||   |||||||||                             ||||||||   |||
Db     5663 CCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCC---TCG 5719

Qy       48 AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg 67
                             |||                  |||:::        :::||||||
Db     5720 CCGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGA 5779

Qy       68 LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSer 87
                                  ::::::                       |||   |||
Db     5780 TCGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCT 5839

Qy       88 Ser---------AsnTrpVal--------------------------ProLysAla--- 94
            |||           |||                              |||:::|||
Db     5840 TCCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCA 5899

Qy       95 ------------HisAlaAspProAspThr------------------LeuProTrpSer 104
                        ||||||  |||   :::                   |||    :::
Db     5900 CCTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACC 5959

Qy      105 LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeu 124
                     ||||||        |||          :::|||      :::
Db     5960 ACGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCG 6019

Qy      125 GluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArg 144
                           |||       |||       :::            |||
Db     6020 ACGGCGTACGTGCCCTTC---ACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTC 6076

Qy      145 ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp 164
            |||||||||   |||                 |||:::                |||:::
Db     6077 ACGCTGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCAT 6136

Qy      165 GluProAla---------AsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
               |||:::          :::         |||      |||:::||||||      |||
Db     6137 GCGCCTTCCTTCCCACCCTCCGCAAGGACGCGACGACGCCGAGGCGTTCACCGCCGCGC 6196

Qy      182 SerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProVal 201
             ||||||      |||                             :::         |||
Db     6197 TCGGCGCTCTCCACTCCG------------------------CAGGCATCACACCCGACT 6232

Qy      202 AsnThrGlyLysSerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
                         |||   |||||||||
Db     6233 GGAGCGCTTTCTTCGCCCCCTTCGCTCCAC---------------------------- 6262

Qy      222 AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValPro---Ile 240
                                           |||   :::   :::||||||:::|||
Db     6263 --------------------------GCAAGGTCTCCCTCCCCACCTATGCCTTCCAGC 6295

Qy      241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
            |||||||||                  |||  |||   |||              ||||||:::
Db     6296 GCGAGCGCTTCTGGCCCGACGCCTCCAAGGCACCCGGCG------------CCGACGTCA 6343

Qy      261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
                        |||:::                        :::|||   ||||||||
Db     6344 GCCACCTTGCTCCGCTCGAGGGGGGGCTCTGGCAAGCCATCGAGCGCGGGGACCTCGATG 6403

Qy      281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
               |||:::|||   |||                              :::|||
Db     6404 CGCTCAGCGGTCAGCTCCACGTGGACGGCGACGAGCGGCGCGCCGCGCTCGCCCTGCTCC 6463

Qy      301 AlaProProSer-------------------GlnProArgGluLysArgProValAsp 313
               |||||||||                   :::   |||   :::
Db     6464 TTCCCACCCTCTCGAGCTTTCGCCACGAGCGGCAAGAGCAGAGCACGGTCGACGCCTGGC 6523

Qy      314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys 333
            |||||||||                         ||||||
Db     6524 GCTACCGTA----------------------TCACCTGGAAGCCTCTGACCACCGCCG 6559

Qy      334 SerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThr 353
               |||           |||        ||||||               ::::::|||
Db     6560 AAACACCCGCCGACCTCGCCGGCACCTGGCTCGTCGTCGTGCCGGCCGCTCTGGACGACG 6619

Qy      354 TyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSer 373
                  |||   |||   |||                        |||   ||||||
Db     6620 ---ACGCGCTCCCCTCCGCGCTCA-----------------------CCGAGGCGCTCA 6652

Qy      374 AsnGlyLeuAlaArgSerGluGlu--------------------------SerAlaArg 384
               |||   ||||||:::                                  :::||||||
Db     6653 CCCGGCGCGGCGCGCGCGTCCTCGCCTTGCGCCTGAGCCAGGCCCACCTGGACCGCGAGG 6712

Qy      385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys 404
                          :::|||   :::   |||||||||                |||:::
Db     6713 CTC--------------TCGCCGAGCATCTGCGCCAGGCTTGCGCCGAGACCGCCCCGA 6757

Qy      405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424
            |||                   :::|||                  ::::::
Db     6758 TTCGCGGCGTGCTCTCGCTCCTCGCCCTCGACGAGCGCC------CCCTCGCAGACCGTC 6811

Qy      425 MetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrp 444
            :::   |||      |||                  |||   |||||       :::
Db     6812 CTGCCCTGCCCGCCGGACTCGCCCTCTCGCTTTCTCTCGCTCAAGCCCTCGGCGACCTCG 6871

Qy      445 SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg 464
            ::::::   |||                                ||||||
Db     6872 ACCTCGAGGCGCCCTTGTGGTTCTTCACGCGCGGCGCCGTCTCCATTGGACACTCTGACC 6931

Qy      465 GlyLysProPheProGluLysPro------------------------------------ 472
               |||   |||   :::|||
Db     6932 CCCTCGCCCATCCCGCCCAGGCCATGACCTGGGGCTTGGGCCGCGTCATCGGCCTCGAGC 6991

Qy      473 ---------ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer------ 487
                           :::|||      |||:::      |||:::         :::
Db     6992 ACCCCGACCGGTGGGGAGGTCTCGTCGACGTCTGCGCTGGGGTCGACGAGAGCGCCGTGG 7051

Qy      488 -----------------LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
                                  ||||||         ::::::      |||
Db     7052 GCCGCTTGCTGCCGGCCCTCGCCGAGCGCCACGACGAAGACCAGCTCGCTCTCCGCCCGG 7111

Qy      502 ProAspAsn---------SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu 518
            ||||||:::         :::      :::||||||            ::::::|||
Db     7112 CCGGACTCTACGCTCGCCGCATCGTCCGCGCCCCGCTCGGCGATGCGCCTCCCG------ 7165

Qy      519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIle 538
                              |||   |||   |||
Db     7166 -----------------CGCGCGACTTCACGC------------------------- 7180

Qy      539 LeuIleProSerProThrSerThrSerProSerValPro---ProValValSerLysGln 557
                  |||         :::||||||:::   |||   |||:::
Db     7181 ------CCGGAGGCACCATTCTCATCACCGGCGGCACCGGCGCCATTGGCGCTCACGTCG 7234
```

```
Qy        558 Pro---GlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
              |||   |||::::::|||||     :::         :::
Db       7235 CCCGATGGCTCGCTCGAAGAGGCGCTCAGCACCTCGTCCTCATCA-------------- 7279

Qy        577 GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596

Db       7279 ------------------------------------------------------------ 7279

Qy        597 PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla 616
                          ||||||    |||:::|||:::
Db       7280 ---------GCCGCCGAGGCGCCGAGGCCCCTG-------------------------- 7303

Qy        617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
                                 |||       -     |||    |||        |||
Db       7304 ------------------GCGCCTCGGAGCTCCACGACGAGCTCTCGGCCCTCGGCGCGC 7345

Qy        637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln 656
                 |||      |||                     :::          |||
Db       7346 GCACCACCCTCGCCGCGTGCGATGTCGCCGACCGGAATGCTGTCGCCACGCTTCTTGAGC 7405

Qy        657 LeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle------GlnLeu 674
                                 :::                     :::|||
Db       7406 AGCTCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGAGCGGCATCGAACACC 7465

Qy        675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis------------------ 688
                 |||      |||      :::      :::   |||
Db       7466 ACGCTCCGCTCGACGCCACCTCTTTCAGGGATCTCGCCGAGGTTGTCTCCGGCAAGGTCG 7525

Qy        689 GlnIleArgHisHisLysThrAspLeuLeuSer------ProSerGlyThrGlyGluVal 706
              :::::::::    |||      ||||||||        |||               :::
Db       7526 AAGGTGCAAAG-CACCTCCACGACCTGCTCGGCTCTCGACCCCTCGACGCCTTTGTTCTC 7584

Qy        707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg--------- 723
                 :::   |||   ||||||:::   |||      |||:::|||:::|||
Db       7585 TTTTCGTCCGGCGC-GGCCGTCTGGGGCGGCGGACAGCAAGGCGGCTACGCGGCCGCAAA 7643

Qy        724 --------------------------------------------------AspValLeu 726
                                                                ||||||
Db       7644 CGCCTTCCTCGACGCCCTTGCCGAGCATCGGCGCAGCGCTGGATTGACAGCGACGTCGGT 7703

Qy        727 GlyLeuGlyAspVal-----------------------------------------Pro 732
              |||||||||   |||                                         |||
Db       7704 GGCCTGGGGCGCGTGGGGCGGCGGCGGCATGGCCACCGATCAGGCGGCAGCCCACCTCCA 7763

Qy        733 SerAlaValGln-GlyArgPheGlyArgAlaLysGlyMet-----------TrpValIl 748
              :::|||           |||            :::      :::              ::
Db       7764 ACAGCGCGGTCTGTCGCGGATGGCCCCCTCGCTTGCCCTGGCGGCGCTCGCGCTGGCTCT 7823

Qy        748 eAspValAspAspThr-----------GlyAspGluAspTrpIleGluThrTyrProSe 764
              ::::   |||:::|||            |||   |||||                 |||||
Db       7824 GGAGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTC 7883

Qy        764 r--------------GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
              |                |||               |||   :::   :::   ||||||
Db       7884 GTTCAGCGCCGCTCGCCCCCGCCCGCTCCTGCGCGAT---TTGCCCGAGGCGCAGCGCGC 7940

Qy        779 rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLe 799
                ||||||             ||||||
Db       7941 TCTCGAGACCAGCGAAGGCGCGTCCTCCGAGCATGGCCCGGCC----------------- 7983

Qy        799 uProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspAr 819
                |||   |||   |||:::   |||     :::   :::         |||          ||
Db       7984 -CCCGACCTCCTCGACAAGCTCCGGAGCCGCTCGGAGAGCGAGCAGCTT---------CG 8033

Qy        819 gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
              ||||:::      |||   |||         :::               |||            |||
Db       8034 TCTGCTCGTCTCGCTGGTGCGCCACGAGACGGCCCTCGTCCTCGGCCACGAAGGCGCCTC 8093

Qy        839 ----------------------------------ValGluPheArgGlnTrp------- 844
                                                |||      |||   |||
Db       8094 CCATGTCGACCCCGACAAGGGCTTCCTCGATCTCGGTCTCGATTCGCTCATGGCCGTCGA 8153

Qy        844 ------------------------------------------------------------ 844

Db       8154 GCTTCGCCGGCGCTTGCAACAGGCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGA 8213

Qy        845 -----------ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV 861
                         :::    ||||||    ||||||    |||||||||
Db       8214 CCATCCCTCTCCTCATCGAGTCGCGCTCTTCTTGCGCGACTCGC---------TCGCCCA 8264

Qy        861 alProPheLeuAlaGlyLeuProAspSer 870
                |||         |||   |||   |||
Db       8265 CGCCCTCGGCACGAGGCTCTCCGTCGAGC 8293
```

RESULT 6
US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-457-342-6

Alignment Scores:

Pred. No.:	0.000564	Length:	28958
Score:	178.50	Matches:	213
Percent Similarity:	29.87%	Conservative:	89
Best Local Similarity:	21.07%	Mismatches:	399
Query Match:	2.41%	Indels:	312

```
DB:            1            Gaps:         42
US-09-913-878A-2 (1-1402) x US-08-457-342-6 (1-28958)

Qy        38 ProHisArgArgLys-----------------------------GluLeuAlaGluSer 47
Db      5663 CCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCC---TCG 5719

Qy        48 AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg 67
Db      5720 CCGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGA 5779

Qy        68 LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSer 87
Db      5780 TCGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCT 5839

Qy        88 Ser---------AsnTrpVal--------------------------ProLysAla--- 94
Db      5840 TCCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCA 5899

Qy        95 ------------HisAlaAspProAspThr------------------LeuProTrpSer 104
Db      5900 CCTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACC 5959

Qy       105 LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeu 124
Db      5960 ACGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCG 6019

Qy       125 GluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArg 144
Db      6020 ACGGCGTACGTGCCCTTC---ACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTC 6076

Qy       145 ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp 164
Db      6077 ACGCTGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCAT 6136

Qy       165 GluProAla---------AsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
Db      6137 GCGCCTTCCTTCCCACCCTCCGCAAGGGACGCGACGACGCCGAGGCGTTCACCGCCGCGC 6196

Qy       182 SerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProVal 201
Db      6197 TCGGCGCTCTCCACTCCG-----------------------CAGGCATCACACCCGACT 6232

Qy       202 AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
Db      6233 GGAGCGCTTTCTTCGCCCCCTTCGCTCCAC----------------------------- 6262

Qy       222 AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValPro---Ile 240
Db      6263 ---------------------------GCAAGGTCTCCCTCCCCACCTATGCCTTCCAGC 6295

Qy       241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
Db      6296 GCGAGCGCTTCTGGCCCGACGCCTCCAAGGCACCCGGCG------------CCGACGTCA 6343

Qy       261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
Db      6344 GCCACCTTGCTCCGCTCGAGGGGGGGCTCTGGCAAGCCATCGAGCGCGGGGACCTCGATG 6403

Qy       281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
Db      6404 CGCTCAGCGGTCAGCTCCACGTGGACGGCGACGAGCGGCGCGCCGCGCTCGCCCTGCTCC 6463

Qy       301 AlaProProSer--------------------GlnProArgGluLysArgProValAsp 313
Db      6464 TTCCCACCCTCTCGAGCTTTCGCCACGAGCGGCAAGAGCAGAGCACGGTCGACGCCTGGC 6523

Qy       314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys 333
Db      6524 GCTACCGTA-----------------------TCACCTGGAAGCCTCTGACCACCGCCG 6559

Qy       334 SerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThr 353
Db      6560 AAACACCCGCCGACCTCGCCGGCACCTGGCTCGTCGTCGTGCCGGCCGCTCTGGACGACG 6619

Qy       354 TyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSer 373
Db      6620 ---ACGCGCTCCCCTCCGCGCTCA-----------------------CCGAGGCGCTCA 6652

Qy       374 AsnGlyLeuAlaArgSerGluGlu--------------------------SerAlaArg 384
Db      6653 CCCGGCGCGGCGCGCGCGTCCTCGCCTTGCGCCTGAGCCAGGCCCACCTGGACCGCGAGG 6712

Qy       385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys 404
Db      6713 CTC---------------TCGCCGAGCATCTGCGCCAGGCTTGCGCCGAGACCGCCCCGA 6757

Qy       405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424
Db      6758 TTCGCGGCGTGCTCTCGCTCCTCGCCCTCGACGAGCGCC------CCCTCGCAGACCGTC 6811

Qy       425 MetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrp 444
Db      6812 CTGCCCTGCCCGCCGGACTCGCCCTCTCGCTTTCTCTCGCTCAAGCCCTCGGCGACCTCG 6871

Qy       445 SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg 464
Db      6872 ACCTCGAGGCGCCCTTGTGGTTCTTCACGCGCGGCGCCGTCTCCATTGGACACTCTGACC 6931

Qy       465 GlyLysProPheProGluLysPro---------------------------------- 472
Db      6932 CCCTCGCCCATCCCGCCCAGGCCATGACCTGGGGCTTGGGCCGCGTCATCGGCCTCGAGC 6991

Qy       473 ---------ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer------ 487
Db      6992 ACCCCGACCGGTGGGGAGGTCTCGTCGACGTCTGCGCTGGGGTCGACGAGAGCGCCGTGG 7051

Qy       488 -----------------LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
Db      7052 GCCGCTTGCTGCCGGCCCTCGCCGAGCGCCACGACGAAGACCAGCTCGCTCTCCGCCCGG 7111

Qy       502 ProAspAsn---------SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu 518
Db      7112 CCGGACTCTACGCTCGCCGCATCGTCCGCGCCCCGCTCGGCGATGCGCCTCCCG------ 7165

Qy       519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIle 538
Db      7166 -----------------CGCGCGACTTCACGC------------------------- 7180

Qy       539 LeuIleProSerProThrSerThrSerProSerValPro---ProValValSerLysGln 557
Db      7181 ------CCGGAGGCACCATTCTCATCACCGGCGGCACCGGCGCCATTGGCGCTCACGTCG 7234

Qy       558 Pro---GlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
Db      7235 CCCGATGGCTCGCTCGAAGAGGCGCTCAGCACCTCGTCCTCATCA-------------- 7279

Qy       577 GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596
Db      7279 ------------------------------------------------------------ 7279

Qy       597 PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla 616
Db      7280 ---------GCCGCCGAGGCGCCGAGGCCCCTG-------------------------- 7303

Qy       617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
Db      7304 -----------------GCGCCTCGGAGCTCCACGACGAGCTCTCGGCCCTCGGCGCGC 7345

Qy       637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln 656
Db      7346 GCACCACCCTCGCCGCGTGCGATGTCGCCGACCGGAATGCTGTCGCCACGCTTCTTGAGC 7405

Qy       657 LeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle------GlnLeu 674
```

```
Db       7406 AGCTCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGAGCGGCATCGAACACC 7465

Qy        675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis----------------- 688
                   |||    |||        :::    :::  |||
Db       7466 ACGCTCCGCTCGACGCCACCTCTTTCAGGGATCTCGCCGAGGTTGTCTCCGGCAAGGTCG 7525

Qy        689 GlnIleArgHisHisLysThrAspLeuLeuSer------ProSerGlyThrGlyGluVal 706
              :::::::::  |||      ||||||||        |||                 :::
Db       7526 AAGGTGCAAAG-CACCTCCACGACCTGCTCGGCTCTCGACCCCTCGACGCCTTTGTTCTC 7584

Qy        707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg--------- 723
                 :::  |||   ||||||:::   |||      |||:::|||:::|||
Db       7585 TTTTCGTCCGGCGC-GGCCGTCTGGGGCGGCGGACAGCAAGGCGGCTACGCGGCCGCAAA 7643

Qy        724 ---------------------------------------------------AspValLeu 726
                                                                 ||||||
Db       7644 CGCCTTCCTCGACGCCCTTGCCGAGCATCGGCGCAGCGCTGGATTGACAGCGACGTCGGT 7703

Qy        727 GlyLeuGlyAspVal------------------------------------------Pro 732
              ||||||||   |||                                          |||
Db       7704 GGCCTGGGGCGCGTGGGGCGGCGGCGGCATGGCCACCGATCAGGCGGCAGCCCACCTCCA 7763

Qy        733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet------------TrpValIl 748
              :::|||      |||         :::      :::                  ::
Db       7764 ACAGCGCGGTCTGTCGCGGATGGCCCCCTCGCTTGCCCTGGCGGCGCTCGCGCTGGCTCT 7823

Qy        748 eAspValAspAspThr-----------GlyAspGluAspTrpIleGluThrTyrProSe 764
              ::::   |||:::|||            |||   |||||          |||||
Db       7824 GGAGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTC 7883

Qy        764 r---------------GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
              |               |||         |||   :::   :::   |||||
Db       7884 GTTCAGCGCCGCTCGCCCCCGCCCGCTCCTGCGCGAT---TTGCCCGAGGCGCAGCGCGC 7940

Qy        779 rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLe 799
                |||||          ||||||
Db       7941 TCTCGAGACCAGCGAAGGCGCGTCCTCCGAGCATGGCCCGGCC----------------- 7983

Qy        799 uProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspAr 819
               |||   |||   |||:::  |||   :::   :::      |||         ||
Db       7984 -CCCGACCTCCTCGACAAGCTCCGGAGCCGCTCGGAGAGCGAGCAGCTT---------CG 8033

Qy        819 gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
              ||||:::       |||   |||         :::         |||         |||
Db       8034 TCTGCTCGTCTCGCTGGTGCGCCACGAGACGGCCCTCGTCCTCGGCCACGAAGGCGCCTC 8093

Qy        839 ----------------------------------ValGluPheArgGlnTrp------- 844
                                                |||      |||   |||
Db       8094 CCATGTCGACCCCGACAAGGGCTTCCTCGATCTCGGTCTCGATTCGCTCATGGCCGTCGA 8153

Qy        844 ------------------------------------------------------------ 844

Db       8154 GCTTCGCCGGCGCTTGCAACAGGCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGA 8213

Qy        845 ----------ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV 861
                        :::  ||||||  ||||||  |||||||||
Db       8214 CCATCCCTCTCCTCATCGAGTCGCGCTCTTCTTGCGCGACTCGC---------TCGCCCA 8264

Qy        861 alProPheLeuAlaGlyLeuProAspSer 870
                |||       |||    |||    |||
Db       8265 CGCCCTCGGCACGAGGCTCTCCGTCGAGC 8293
```

RESULT 7
US-08-457-646A-6
; Sequence 6, Application US/08457646A
; Patent No. 5679560
; GENERAL INFORMATION:
;   APPLICANT: Schupp, Thomas
;   APPLICANT: Ligon, James M.
;   APPLICANT: Beck, James Joseph
;   APPLICANT: Hill, Dwight Steven
;   APPLICANT: Ryals, John Andrew
;   APPLICANT: Gaffney, Thomas Deane
;   APPLICANT: Lam, Stephen Ting
;   APPLICANT: Hammer, Phillip E.
;   APPLICANT: Uknes, Scott Joseph
;   TITLE OF INVENTION: Genes for the synthesis of
;   TITLE OF INVENTION: antipathogenic substances
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ciba-Geigy Corporation
;   STREET: 7 Skyline Drive
;   CITY: Hawthorne
;   STATE: NY
;   COUNTRY: USA
;   ZIP: 10532
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/457,646A
;   FILING DATE: 01-JUN-1995
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/457,205
;   FILING DATE: 01-JUN-1995
;   APPLICATION NUMBER: 08/258,261
;   FILING DATE: 08-Jun-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Elmer, James Scott
;   REGISTRATION NUMBER: 36,129
;   REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 919-541-8614
;   TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 28958 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
US-08-457-646A-6

Alignment Scores:

Pred. No.:	0.000564	Length:	28958
Score:	178.50	Matches:	213
Percent Similarity:	29.87%	Conservative:	89
Best Local Similarity:	21.07%	Mismatches:	399
Query Match:	2.41%	Indels:	312
DB:	1	Gaps:	42

US-09-913-878A-2 (1-1402) x US-08-457-646A-6 (1-28958)

```
Qy         38 ProHisArgArgLys-----------------------------GluLeuAlaGluSer 47
              |||   |||||||||                             |||||||||   |||
Db       5663 CCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCC---TCG 5719

Qy         48 AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg 67
                       |||                  |||:::          :::||||||
Db       5720 CCGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGA 5779

Qy         68 LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSer 87
                                ::::::                  |||   |||
Db       5780 TCGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCT 5839

Qy         88 Ser---------AsnTrpVal--------------------------ProLysAla--- 94
              |||         |||                                |||:::|||
Db       5840 TCCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCA 5899

Qy         95 ------------HisAlaAspProAspThr------------------LeuProTrpSer 104
                          ||||||   |||   :::                  |||   :::
```

```
Db   5900 CCTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACC 5959

Qy    105 LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeu 124
Db   5960 ACGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCG 6019

Qy    125 GluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArg 144
Db   6020 ACGGCGTACGTGCCCTTC---ACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTC 6076

Qy    145 ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp 164
Db   6077 ACGCTGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCAT 6136

Qy    165 GluProAla---------AsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
Db   6137 GCGCCTTCCTTCCCACCCTCCGCAAGGGACGCGACGACGCCGAGGCGTTCACCGCCGCGC 6196

Qy    182 SerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProVal 201
Db   6197 TCGGCGCTCTCCACTCCG-----------------------CAGGCATCACACCCGACT 6232

Qy    202 AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
Db   6233 GGAGCGCTTTCTTCGCCCCCTTCGCTCCAC---------------------------- 6262

Qy    222 AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValPro---Ile 240
Db   6263 --------------------------GCAAGGTCTCCCTCCCCACCTATGCCTTCCAGC 6295

Qy    241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
Db   6296 GCGAGCGCTTCTGGCCCGACGCCTCCAAGGCACCCGGCG-----------CCGACGTCA 6343

Qy    261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
Db   6344 GCCACCTTGCTCCGCTCGAGGGGGGGCTCTGGCAAGCCATCGAGCGCGGGGACCTCGATG 6403

Qy    281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
Db   6404 CGCTCAGCGGTCAGCTCCACGTGGACGGCGACGAGCGGCGCGCCGCGCTCGCCCTGCTCC 6463

Qy    301 AlaProProSer--------------------GlnProArgGluLysArgProValAsp 313
Db   6464 TTCCCACCCTCTCGAGCTTTCGCCACGAGCGGCAAGAGCAGAGCACGGTCGACGCCTGGC 6523

Qy    314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys 333
Db   6524 GCTACCGTA-----------------------TCACCTGGAAGCCTCTGACCACCGCCG 6559

Qy    334 SerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThr 353
Db   6560 AAACACCCGCCGACCTCGCCGGCACCTGGCTCGTCGTCGTGCCGGCCGCTCTGGACGACG 6619

Qy    354 TyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSer 373
Db   6620 ---ACGCGCTCCCCTCCGCGCTCA-----------------------CCGAGGCGCTCA 6652

Qy    374 AsnGlyLeuAlaArgSerGluGlu--------------------------SerAlaArg 384
Db   6653 CCCGGCGCGGCGCGCGCGTCCTCGCCTTGCGCCTGAGCCAGGCCCACCTGGACCGCGAGG 6712

Qy    385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys 404
Db   6713 CTC---------------TCGCCGAGCATCTGCGCCAGGCTTGCGCCGAGACCGCCCCGA 6757

Qy    405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424
Db   6758 TTCGCGGCGTGCTCTCGCTCCTCGCCCTCGACGAGCGCC------CCCTCGCAGACCGTC 6811

Qy    425 MetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrp 444
Db   6812 CTGCCCTGCCCGCCGGACTCGCCCTCTCGCTTTCTCTCGCTCAAGCCCTCGGCGACCTCG 6871

Qy    445 SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg 464
Db   6872 ACCTCGAGGCGCCCTTGTGGTTCTTCACGCGCGGCGCCGTCTCCATTGGACACTCTGACC 6931

Qy    465 GlyLysProPheProGluLysPro----------------------------------- 472
Db   6932 CCCTCGCCCATCCCGCCCAGGCCATGACCTGGGGCTTGGGCCGCGTCATCGGCCTCGAGC 6991

Qy    473 ---------ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer------ 487
Db   6992 ACCCCGACCGGTGGGGAGGTCTCGTCGACGTCTGCGCTGGGGTCGACGAGAGCGCCGTGG 7051

Qy    488 -----------------LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
Db   7052 GCCGCTTGCTGCCGGCCCTCGCCGAGCGCCACGACGAAGACCAGCTCGCTCTCCGCCCGG 7111

Qy    502 ProAspAsn---------SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu 518
Db   7112 CCGGACTCTACGCTCGCCGCATCGTCCGCGCCCCGCTCGGCGATGCGCCTCCCG------ 7165

Qy    519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIle 538
Db   7166 -----------------CGCGCGACTTCACGC--------------------------- 7180

Qy    539 LeuIleProSerProThrSerThrSerProSerValPro---ProValValSerLysGln 557
Db   7181 ------CCGGAGGCACCATTCTCATCACCGGCGGCACCGGCGCCATTGGCGCTCACGTCG 7234

Qy    558 Pro---GlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
Db   7235 CCCGATGGCTCGCTCGAAGAGGCGCTCAGCACCTCGTCCTCATCA-------------- 7279

Qy    577 GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596
Db   7279 ------------------------------------------------------------ 7279

Qy    597 PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla 616
Db   7280 ---------GCCGCCGAGGCGCCGAGGCCCCTG-------------------------- 7303

Qy    617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
Db   7304 -----------------GCGCCTCGGAGCTCCACGACGAGCTCTCGGCCCTCGGCGCGC 7345

Qy    637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln 656
Db   7346 GCACCACCCTCGCCGCGTGCGATGTCGCCGACCGGAATGCTGTCGCCACGCTTCTTGAGC 7405

Qy    657 LeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle------GlnLeu 674
Db   7406 AGCTCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGAGCGGCATCGAACACC 7465

Qy    675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis------------------ 688
Db   7466 ACGCTCCGCTCGACGCCACCTCTTTCAGGGATCTCGCCGAGGTTGTCTCCGGCAAGGTCG 7525

Qy    689 GlnIleArgHisHisLysThrAspLeuLeuSer------ProSerGlyThrGlyGluVal 706
Db   7526 AAGGTGCAAAG-CACCTCCACGACCTGCTCGGCTCTCGACCCCTCGACGCCTTTGTTCTC 7584

Qy    707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg--------- 723
Db   7585 TTTTCGTCCGGCGC-GGCCGTCTGGGGCGGCGGACAGCAAGGCGGCTACGCGGCCGCAAA 7643

Qy    724 ------------------------------------------------AspValLeu 726
Db   7644 CGCCTTCCTCGACGCCCTTGCCGAGCATCGGCGCAGCGCTGGATTGACAGCGACGTCGGT 7703

Qy    727 GlyLeuGlyAspVal-------------------------------------------Pro 732
Db   7704 GGCCTGGGGCGCGTGGGGCGGCGGCGGCATGGCCACCGATCAGGCGGCAGCCCACCTCCA 7763
```

```
Qy        733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet------------TrpValIl 748
              :::|||         |||         :::    :::                  ::
Db       7764 ACAGCGCGGTCTGTCGCGGATGGCCCCCTCGCTTGCCCTGGCGGCGCTCGCGCTGGCTCT 7823

Qy        748 eAspValAspAspThr------------GlyAspGluAspTrpIleGluThrTyrProSe 764
              ::::   |||:::|||            |||   ||||||             |||||
Db       7824 GGAGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTC 7883

Qy        764 r---------------GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
              |               |||          |||   :::   :::   ||||||
Db       7884 GTTCAGCGCCGCTCGCCCCCGCCCGCTCCTGCGCGAT---TTGCCCGAGGCGCAGCGCGC 7940

Qy        779 rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLe 799
                ||||||            ||||||
Db       7941 TCTCGAGACCAGCGAAGGCGCGTCCTCCGAGCATGGCCCGGCC----------------- 7983

Qy        799 uProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspAr 819
                |||   |||   |||:::   |||   :::   :::      |||            ||
Db       7984 -CCCGACCTCCTCGACAAGCTCCGGAGCCGCTCGGAGAGCGAGCAGCTT---------CG 8033

Qy        819 gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
              ||||:::      |||   |||      :::           |||         |||
Db       8034 TCTGCTCGTCTCGCTGGTGCGCCACGAGACGGCCCTCGTCCTCGGCCACGAAGGCGCCTC 8093

Qy        839 ----------------------------------ValGluPheArgGlnTrp------- 844
                                                  |||   |||   |||
Db       8094 CCATGTCGACCCCGACAAGGGCTTCCTCGATCTCGGTCTCGATTCGCTCATGGCCGTCGA 8153

Qy        844 ------------------------------------------------------------ 844

Db       8154 GCTTCGCCGGCGCTTGCAACAGGCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGA 8213

Qy        845 ----------ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV 861
                        :::   ||||||   ||||||   |||||||||
Db       8214 CCATCCCTCTCCTCATCGAGTCGCGCTCTTCTTGCGCGACTCGC---------TCGCCCA 8264

Qy        861 alProPheLeuAlaGlyLeuProAspSer 870
                |||      |||   |||   |||
Db       8265 CGCCCTCGGCACGAGGCTCTCCGTCGAGC 8293
```

RESULT 8
US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,076A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-458-076A-6

```
Alignment Scores:
Pred. No.:               0.000564     Length:        28958
Score:                   178.50       Matches:       213
Percent Similarity:      29.87%       Conservative:  89
Best Local Similarity:   21.07%       Mismatches:    399
Query Match:             2.41%        Indels:        312
DB:                      1            Gaps:          42

US-09-913-878A-2 (1-1402) x US-08-458-076A-6 (1-28958)

Qy         38 ProHisArgArgLys-----------------------------GluLeuAlaGluSer 47
              |||   |||||||||                             |||||||||   |||
Db       5663 CCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCC---TCG 5719

Qy         48 AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg 67
                          |||                     |||:::            :::||||||
Db       5720 CCGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGA 5779

Qy         68 LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSer 87
                                    ::::::                  |||   |||
Db       5780 TCGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCT 5839

Qy         88 Ser---------AsnTrpVal--------------------------ProLysAla--- 94
              |||          |||                               |||:::|||
Db       5840 TCCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCA 5899

Qy         95 ------------HisAlaAspProAspThr------------------LeuProTrpSer 104
                          ||||||   |||   :::                    |||   :::
Db       5900 CCTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACC 5959

Qy        105 LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeu 124
                    ||||||   |||               :::|||      :::
Db       5960 ACGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCG 6019

Qy        125 GluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArg 144
                             |||   |||      :::                :::   |||
Db       6020 ACGGCGTACGTGCCCTTC---ACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTC 6076

Qy        145 ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp 164
              |||||||||      |||                  |||:::               |||:::
Db       6077 ACGCTGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCAT 6136

Qy        165 GluProAla---------AsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
                 |||:::          :::            |||   |||:::||||||      |||
Db       6137 GCGCCTTCCTTCCCACCCTCCGCAAGGGACGCGACGACGCCGAGGCGTTCACCGCCGCGC 6196

Qy        182 SerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProVal 201
              ||||||         |||                              :::   |||
Db       6197 TCGGCGCTCTCCACTCCG-----------------------CAGGCATCACACCCGACT 6232
```

```
Qy        202 AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
                          |||       |||||||||
Db       6233 GGAGCGCTTTCTTCGCCCCCTTCGCTCCAC------------------------------ 6262

Qy        222 AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValPro---Ile 240
                                              |||   :::   :::|||||||:::|||
Db       6263 ---------------------------GCAAGGTCTCCCTCCCCACCTATGCCTTCCAGC 6295

Qy        241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
              |||||||||               |||   |||         |||         ||||||:::
Db       6296 GCGAGCGCTTCTGGCCCGACGCCTCCAAGGCACCCGGCG------------CCGACGTCA 6343

Qy        261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
                             |||:::                           :::|||   |||||||||
Db       6344 GCCACCTTGCTCCGCTCGAGGGGGGGCTCTGGCAAGCCATCGAGCGCGGGGACCTCGATG 6403

Qy        281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
                 |||:::|||   |||                                 :::|||
Db       6404 CGCTCAGCGGTCAGCTCCACGTGGACGGCGACGAGCGGCGCGCCGCGCTCGCCCTGCTCC 6463

Qy        301 AlaProProSer--------------------GlnProArgGluLysArgProValAsp 313
                 |||||||||                        :::   |||   :::
Db       6464 TTCCCACCCTCTCGAGCTTTCGCCACGAGCGGCAAGAGCAGAGCACGGTCGACGCCTGGC 6523

Qy        314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys 333
              |||||||||                                 ||||||
Db       6524 GCTACCGTA------------------------TCACCTGGAAGCCTCTGACCACCGCCG 6559

Qy        334 SerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThr 353
                 |||         |||         ||||||                  ::::::|||
Db       6560 AAACACCCGCCGACCTCGCCGGCACCTGGCTCGTCGTCGTGCCGGCCGCTCTGGACGACG 6619

Qy        354 TyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSer 373
                       |||   |||   |||                           |||   ||||||
Db       6620 ---ACGCGCTCCCCTCCGCGCTCA-----------------------CCGAGGCGCTCA 6652

Qy        374 AsnGlyLeuAlaArgSerGluGlu--------------------------SerAlaArg 384
                 |||   ||||||:::                                    :::||||||
Db       6653 CCCGGCGCGGCGCGCGCGTCCTCGCCTTGCGCCTGAGCCAGGCCCACCTGGACCGCGAGG 6712

Qy        385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys 404
                                   :::|||   :::   |||||||||            :::|||
Db       6713 CTC---------------TCGCCGAGCATCTGCGCCAGGCTTGCGCCGAGACCGCCCCGA 6757

Qy        405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424
              |||                   :::|||         |||         ::::::
Db       6758 TTCGCGGCGTGCTCTCGCTCCTCGCCCTCGACGAGCGCC------CCCTCGCAGACCGTC 6811

Qy        425 MetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrp 444
              :::   |||      |||               |||   ||||||         :::
Db       6812 CTGCCCTGCCCGCCGGACTCGCCCTCTCGCTTTCTCTCGCTCAAGCCCTCGGCGACCTCG 6871

Qy        445 SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg 464
              ::::::   |||                                    ||||||
Db       6872 ACCTCGAGGCGCCCTTGTGGTTCTTCACGCGCGGCGCCGTCTCCATTGGACACTCTGACC 6931

Qy        465 GlyLysProPheProGluLysPro----------------------------------- 472
                    |||   |||   :::|||
Db       6932 CCCTCGCCCATCCCGCCCAGGCCATGACCTGGGGCTTGGGCCGCGTCATCGGCCTCGAGC 6991

Qy        473 ---------ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer------ 487
                          :::|||      |||:::      |||:::      :::
Db       6992 ACCCCGACCGGTGGGGAGGTCTCGTCGACGTCTGCGCTGGGGTCGACGAGAGCGCCGTGG 7051

Qy        488 -----------------LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
                                   ||||||         ::::::   |||
Db       7052 GCCGCTTGCTGCCGGCCCTCGCCGAGCGCCACGACGAAGACCAGCTCGCTCTCCGCCCGG 7111

Qy        502 ProAspAsn---------SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu 518
              ||||||:::               :::   :::||||||         ::::::|||
Db       7112 CCGGACTCTACGCTCGCCGCATCGTCCGCGCCCCGCTCGGCGATGCGCCTCCCG------ 7165

Qy        519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIle 538
                                                  |||   |||   |||
Db       7166 -----------------CGCGCGACTTCACGC-------------------------- 7180

Qy        539 LeuIleProSerProThrSerThrSerProSerValPro---ProValValSerLysGln 557
                    |||               :::||||||:::      |||      |||:::
Db       7181 ------CCGGAGGCACCATTCTCATCACCGGCGGCACCGGCGCCATTGGCGCTCACGTCG 7234

Qy        558 Pro---GlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
              |||   |||::::::|||||||         :::            :::
Db       7235 CCCGATGGCTCGCTCGAAGAGGCGCTCAGCACCTCGTCCTCATCA--------------- 7279

Qy        577 GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596

Db       7279 ------------------------------------------------------------ 7279

Qy        597 PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla 616
                                ||||||   |||:::|||:::
Db       7280 ---------GCCGCCGAGGCGCCGAGGCCCCTG------------------------- 7303

Qy        617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
                                   |||               |||   |||         |||
Db       7304 -----------------GCGCCTCGGAGCTCCACGACGAGCTCTCGGCCCTCGGCGCGC 7345

Qy        637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln 656
                 |||         |||                     :::            |||
Db       7346 GCACCACCCTCGCCGCGTGCGATGTCGCCGACCGGAATGCTGTCGCCACGCTTCTTGAGC 7405

Qy        657 LeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle------GlnLeu 674
                                :::                  :::|||
Db       7406 AGCTCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGAGCGGCATCGAACACC 7465

Qy        675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis------------------ 688
                 |||      |||      :::      :::   |||
Db       7466 ACGCTCCGCTCGACGCCACCTCTTTCAGGGATCTCGCCGAGGTTGTCTCCGGCAAGGTCG 7525

Qy        689 GlnIleArgHisHisLysThrAspLeuLeuSer------ProSerGlyThrGlyGluVal 706
              ::::::::   |||         |||||||||         |||            :::
Db       7526 AAGGTGCAAAG-CACCTCCACGACCTGCTCGGCTCTCGACCCCTCGACGCCTTTGTTCTC 7584

Qy        707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg--------- 723
                 :::   |||   ||||||:::   |||      |||:::|||:::|||
Db       7585 TTTTCGTCCGGCGC-GGCCGTCTGGGGCGGCGGACAGCAAGGCGGCTACGCGGCCGCAAA 7643

Qy        724 -------------------------------------------------AspValLeu 726
                                                               ||||||
Db       7644 CGCCTTCCTCGACGCCCTTGCCGAGCATCGGCGCAGCGCTGGATTGACAGCGACGTCGGT 7703

Qy        727 GlyLeuGlyAspVal-----------------------------------------Pro 732
              |||||||||   |||                                         |||
Db       7704 GGCCTGGGGCGCGTGGGGCGGCGGCGGCATGGCCACCGATCAGGCGGCAGCCCACCTCCA 7763

Qy        733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet-----------TrpValIl 748
              ::::||            |||         :::      :::              ::
Db       7764 ACAGCGCGGTCTGTCGCGGATGGCCCCCTCGCTTGCCCTGGCGGCGCTCGCGCTGGCTCT 7823

Qy        748 eAspValAspAspThr------------GlyAspGluAspTrpIleGluThrTyrProSe 764
               :::   |||:::|||         |||      ||||||         ||||
Db       7824 GGAGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTC 7883

Qy        764 r---------------GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
              |               |||         |||   :::   :::   ||||||
Db       7884 GTTCAGCGCCGCTCGCCCCCGCCCGCTCCTGCGCGAT---TTGCCCGAGGCGCAGCGCGC 7940

Qy        779 rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLe 799
               ||||||            ||||||
Db       7941 TCTCGAGACCAGCGAAGGCGCGTCCTCCGAGCATGGCCCGGCC---------------- 7983

Qy        799 uProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspAr 819
               |||   |||   |||:::   |||   :::   :::         |||            ||
Db       7984 -CCCGACCTCCTCGACAAGCTCCGGAGCCGCTCGGAGAGCGAGCAGCTT---------CG 8033

Qy        819 gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
              ||||:::      |||   |||      :::               |||         |||
```

```
Db       8034 TCTGCTCGTCTCGCTGGTGCGCCACGAGACGGCCCTCGTCCTCGGCCACGAAGGCGCCTC 8093

Qy        839 ----------------------------------ValGluPheArgGlnTrp------- 844
                                                |||      |||   |||
Db       8094 CCATGTCGACCCCGACAAGGGCTTCCTCGATCTCGGTCTCGATTCGCTCATGGCCGTCGA 8153

Qy        844 ------------------------------------------------------------ 844

Db       8154 GCTTCGCCGGCGCTTGCAACAGGCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGA 8213

Qy        845 -----------ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV 861
                         :::  ||||||  ||||||  ||||||||
Db       8214 CCATCCCTCTCCTCATCGAGTCGCGCTCTTCTTGCGCGACTCGC---------TCGCCCA 8264

Qy        861 alProPheLeuAlaGlyLeuProAspSer 870
                 |||         |||   |||   |||
Db       8265 CGCCCTCGGCACGAGGCTCTCCGTCGAGC 8293
```

RESULT 9
US-08-764-233A-4
; Sequence 4, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezanna
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Sorangium cellulosum
; IMMEDIATE SOURCE:
; CLONE: p98/1
US-08-764-233A-4

```
Alignment Scores:
Pred. No.:                 0.000564      Length:          28958
Score:                     178.50        Matches:         213
Percent Similarity:        29.87%        Conservative:    89
Best Local Similarity:     21.07%        Mismatches:      399
Query Match:               2.41%         Indels:          312
DB:                        1             Gaps:            42

US-09-913-878A-2 (1-1402) x US-08-764-233A-4 (1-28958)

Qy         38 ProHisArgArgLys-----------------------------GluLeuAlaGluSer 47
              |||  ||||||||                                |||||||||   |||
Db       5663 CCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCC---TCG 5719

Qy         48 AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg 67
                          |||                     |||:::         :::||||||
Db       5720 CCGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGA 5779

Qy         68 LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSer 87
                                         ::::::                 |||   |||
Db       5780 TCGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCT 5839

Qy         88 Ser---------AsnTrpVal--------------------------ProLysAla--- 94
              |||         |||                                |||:::|||
Db       5840 TCCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCA 5899

Qy         95 ------------HisAlaAspProAspThr------------------LeuProTrpSer 104
                          ||||||  |||   :::                   |||   :::
Db       5900 CCTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACC 5959

Qy        105 LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeu 124
                    ||||||      |||               :::|||      :::
Db       5960 ACGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCG 6019

Qy        125 GluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArg 144
                             |||   |||      :::               :::   |||
Db       6020 ACGGCGTACGTGCCCTTC---ACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTC 6076

Qy        145 ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp 164
              |||||||||   |||                |||:::                |||:::
Db       6077 ACGCTGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCAT 6136

Qy        165 GluProAla---------AsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
                 |||:::         :::         |||      |||:::||||||      |||
Db       6137 GCGCCTTCCTTCCCACCCTCCGCAAGGGACGCGACGACGCCGAGGCGTTCACCGCCGCGC 6196

Qy        182 SerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProVal 201
              ||||||      |||                               :::      |||
Db       6197 TCGGCGCTCTCCACTCCG-----------------------CAGGCATCACACCCGACT 6232

Qy        202 AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
                       |||   |||||||||
Db       6233 GGAGCGCTTTCTTCGCCCCCTTCGCTCCAC------------------------------ 6262

Qy        222 AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValPro---Ile 240
                                      |||   :::   :::|||||||:::|||
Db       6263 --------------------------GCAAGGTCTCCCTCCCCACCTATGCCTTCCAGC 6295

Qy        241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
              |||||||||               |||   |||      |||         ||||||:::
Db       6296 GCGAGCGCTTCTGGCCCGACGCCTCCAAGGCACCCGGCG------------CCGACGTCA 6343

Qy        261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
                          |||:::                    :::|||   |||||||||
Db       6344 GCCACCTTGCTCCGCTCGAGGGGGGGCTCTGGCAAGCCATCGAGCGCGGGGACCTCGATG 6403

Qy        281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
                 |||:::|||   |||                        :::|||
Db       6404 CGCTCAGCGGTCAGCTCCACGTGGACGGCGACGAGCGGCGCGCCGCGCTCGCCCTGCTCC 6463

Qy        301 AlaProProSer--------------------GlnProArgGluLysArgProValAsp 313
                 |||||||||                    :::   |||    :::
Db       6464 TTCCCACCCTCTCGAGCTTTCGCCACGAGCGGCAAGAGCAGAGCACGGTCGACGCCTGGC 6523
```

```
Qy        314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys 333
              |||||||||                       ||||||
Db       6524 GCTACCGTA-----------------------TCACCTGGAAGCCTCTGACCACCGCCG 6559

Qy        334 SerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThr 353
                |||        |||       ||||||                :::::|||
Db       6560 AAACACCCGCCGACCTCGCCGGCACCTGGCTCGTCGTCGTGCCGGCCGCTCTGGACGACG 6619

Qy        354 TyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSer 373
                       |||   |||   |||                      |||   ||||||
Db       6620 ---ACGCGCTCCCCTCCGCGCTCA-----------------------CCGAGGCGCTCA 6652

Qy        374 AsnGlyLeuAlaArgSerGluGlu--------------------------SerAlaArg 384
                 |||   ||||||:::                               :::||||||
Db       6653 CCCGGCGCGGCGCGCGTCCTCGCCTTGCGCCTGAGCCAGGCCCACCTGGACCGCGAGG 6712

Qy        385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys 404
                                    :::|||   :::   |||||||||          |||:::
Db       6713 CTC---------------TCGCCGAGCATCTGCGCCAGGCTTGCGCCGAGACCGCCCCGA 6757

Qy        405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424
              |||                   :::|||         |||      ::::::
Db       6758 TTCGCGGCGTGCTCTCGCTCCTCGCCCTCGACGAGCGCC------CCCTCGCAGACCGTC 6811

Qy        425 MetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrp 444
              :::    |||      |||                |||   ||||||       :::
Db       6812 CTGCCCTGCCCGCCGGACTCGCCCTCTCGCTTTCTCTCGCTCAAGCCCTCGGCGACCTCG 6871

Qy        445 SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg 464
              ::::::   |||                                  ||||||
Db       6872 ACCTCGAGGCGCCCTTGTGGTTCTTCACGCGCGGCGCCGTCTCCATTGGACACTCTGACC 6931

Qy        465 GlyLysProPheProGluLysPro----------------------------------- 472
                    |||   |||   :::|||
Db       6932 CCCTCGCCCATCCCGCCCAGGCCATGACCTGGGGCTTGGGCCGCGTCATCGGCCTCGAGC 6991

Qy        473 ---------ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer------ 487
                               :::|||       |||:::      |||:::   :::
Db       6992 ACCCCGACCGGTGGGGAGGTCTCGTCGACGTCTGCGCTGGGGTCGACGAGAGCGCCGTGG 7051

Qy        488 -----------------LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
                                         ||||||        ::::::   |||
Db       7052 GCCGCTTGCTGCCGGCCCTCGCCGAGCGCCACGACGAAGACCAGCTCGCTCTCCGCCCGG 7111

Qy        502 ProAspAsn---------SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu 518
              ||||||:::          :::      :::||||||      ::::::|||
Db       7112 CCGGACTCTACGCTCGCCGCATCGTCCGCGCCCCGCTCGGCGATGCGCCTCCCG------ 7165

Qy        519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIle 538
                                   |||   |||   |||
Db       7166 -----------------CGCGCGACTTCACGC-------------------------- 7180

Qy        539 LeuIleProSerProThrSerThrSerProSerValPro---ProValValSerLysGln 557
                    |||          :::||||||:::      |||   |||:::
Db       7181 ------CCGGAGGCACCATTCTCATCACCGGCGGCACCGGCGCCATTGGCGCTCACGTCG 7234

Qy        558 Pro---GlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
              |||   |||::::::||||||      :::
Db       7235 CCCGATGGCTCGCTCGAAGAGGCGCTCAGCACCTCGTCCTCATCA--------------- 7279

Qy        577 GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596

Db       7279 ------------------------------------------------------------ 7279

Qy        597 PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla 616
                             ||||||   |||:::|||:::
Db       7280 ---------GCCGCCGAGGCGCCGAGGCCCCTG-------------------------- 7303

Qy        617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
                                   |||         |||   |||         |||
Db       7304 ------------------GCGCCTCGGAGCTCCACGACGAGCTCTCGGCCCTCGGCGCGC 7345

Qy        637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln 656
                 |||      |||                           :::            |||
Db       7346 GCACCACCCTCGCCGCGTGCGATGTCGCCGACCGGAATGCTGTCGCCACGCTTCTTGAGC 7405

Qy        657 LeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle------GlnLeu 674
                             :::                     :::|||
Db       7406 AGCTCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGAGCGGCATCGAACACC 7465

Qy        675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis----------------- 688
                 |||      |||         :::      :::   |||
Db       7466 ACGCTCCGCTCGACGCCACCTCTTTCAGGGATCTCGCCGAGGTTGTCTCCGGCAAGGTCG 7525

Qy        689 GlnIleArgHisHisLysThrAspLeuLeuSer------ProSerGlyThrGlyGluVal 706
              :::::::::   |||      |||||||||         |||               :::
Db       7526 AAGGTGCAAAG-CACCTCCACGACCTGCTCGGCTCTCGACCCCTCGACGCCTTTGTTCTC 7584

Qy        707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg--------- 723
                 :::   |||   ||||||:::   |||      |||:::|||:::|||
Db       7585 TTTTCGTCCGGCGC-GGCCGTCTGGGGCGGCGGACAGCAAGGCGGCTACGCGGCCGCAAA 7643

Qy        724 -------------------------------------------------AspValLeu 726
                                                                ||||||
Db       7644 CGCCTTCCTCGACGCCCTTGCCGAGCATCGGCGCAGCGCTGGATTGACAGCGACGTCGGT 7703

Qy        727 GlyLeuGlyAspVal------------------------------------------Pro 732
              |||||||||   |||                                          |||
Db       7704 GGCCTGGGGCGCGTGGGGCGGCGGCGGCATGGCCACCGATCAGGCGGCAGCCCACCTCCA 7763

Qy        733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet------------TrpValIl 748
              :::|||           |||      :::   :::                 ::
Db       7764 ACAGCGCGGTCTGTCGCGGATGGCCCCCTCGCTTGCCCTGGCGGCGCTCGCGCTGGCTCT 7823

Qy        748 eAspValAspAspThr------------GlyAspGluAspTrpIleGluThrTyrProSe 764
               ::::   |||:::|||             |||   ||||||            |||||
Db       7824 GGAGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTC 7883

Qy        764 r---------------GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
              |                  |||          |||    :::   :::   ||||||
Db       7884 GTTCAGCGCCGCTCGCCCCCGCCCGCTCCTGCGCGAT---TTGCCCGAGGCGCAGCGCGC 7940

Qy        779 rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLe 799
               ||||||            ||||||
Db       7941 TCTCGAGACCAGCGAAGGCGCGTCCTCCGAGCATGGCCCGGCC----------------- 7983

Qy        799 uProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspAr 819
               |||   |||   |||:::   |||   :::   :::      |||            ||
Db       7984 -CCCGACCTCCTCGACAAGCTCCGGAGCCGCTCGGAGAGCGAGCAGCTT---------CG 8033

Qy        819 gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
              ||||:::      |||   |||      :::          |||            |||
Db       8034 TCTGCTCGTCTCGCTGGTGCGCCACGAGACGGCCCTCGTCCTCGGCCACGAAGGCGCCTC 8093

Qy        839 --------------------------------ValGluPheArgGlnTrp------- 844
                                              |||   |||   |||
Db       8094 CCATGTCGACCCCGACAAGGGCTTCCTCGATCTCGGTCTCGATTCGCTCATGGCCGTCGA 8153

Qy        844 ----------------------------------------------------------- 844

Db       8154 GCTTCGCCGGCGCTTGCAACAGGCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGA 8213

Qy        845 -----------ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV 861
                         :::   ||||||   ||||||   |||||||||
Db       8214 CCATCCCTCTCCTCATCGAGTCGCGCTCTTCTTGCGCGACTCGC---------TCGCCCA 8264

Qy        861 alProPheLeuAlaGlyLeuProAspSer 870
               |||         |||   |||   |||
Db       8265 CGCCCTCGGCACGAGGCTCTCCGTCGAGC 8293
```

RESULT 10
US-08-457-335A-6
; Sequence 6, Application US/08457335A
; Patent No. 5723759
; GENERAL INFORMATION:

```
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,335A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-335A-6

Alignment Scores:
Pred. No.:               0.000564      Length:         28958
Score:                   178.50        Matches:        213
Percent Similarity:      29.87%        Conservative:   89
Best Local Similarity:   21.07%        Mismatches:     399
Query Match:             2.41%         Indels:         312
DB:                      1             Gaps:           42

US-09-913-878A-2 (1-1402) x US-08-457-335A-6 (1-28958)

Qy         38 ProHisArgArgLys-----------------------------GluLeuAlaGluSer 47
Db       5663 CCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCC---TCG 5719

Qy         48 AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg 67
Db       5720 CCGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGA 5779

Qy         68 LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSer 87
Db       5780 TCGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCT 5839

Qy         88 Ser---------AsnTrpVal--------------------------ProLysAla--- 94
Db       5840 TCCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCA 5899

Qy         95 ------------HisAlaAspProAspThr------------------LeuProTrpSer 104
Db       5900 CCTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACC 5959

Qy        105 LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeu 124
Db       5960 ACGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCG 6019

Qy        125 GluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArg 144
Db       6020 ACGGCGTACGTGCCCTTC---ACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTC 6076

Qy        145 ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp 164
Db       6077 ACGCTGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCAT 6136

Qy        165 GluProAla---------AsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
Db       6137 GCGCCTTCCTTCCCACCCTCCGCAAGGGACGCGACGACGCCGAGGCGTTCACCGCCGCGC 6196

Qy        182 SerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProVal 201
Db       6197 TCGGCGCTCTCCACTCCG-----------------------CAGGCATCACACCCGACT 6232

Qy        202 AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
Db       6233 GGAGCGCTTTCTTCGCCCCCTTCGCTCCAC----------------------------- 6262

Qy        222 AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValPro---Ile 240
Db       6263 --------------------------GCAAGGTCTCCCTCCCCACCTATGCCTTCCAGC 6295

Qy        241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
Db       6296 GCGAGCGCTTCTGGCCCGACGCCTCCAAGGCACCCGGCG------------CCGACGTCA 6343

Qy        261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
Db       6344 GCCACCTTGCTCCGCTCGAGGGGGGGCTCTGGCAAGCCATCGAGCGCGGGGACCTCGATG 6403

Qy        281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
Db       6404 CGCTCAGCGGTCAGCTCCACGTGGACGGCGACGAGCGGCGCGCCGCGCTCGCCCTGCTCC 6463

Qy        301 AlaProProSer--------------------GlnProArgGluLysArgProValAsp 313
Db       6464 TTCCCACCCTCTCGAGCTTTCGCCACGAGCGGCAAGAGCAGAGCACGGTCGACGCCTGGC 6523

Qy        314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys 333
Db       6524 GCTACCGTA-----------------------TCACCTGGAAGCCTCTGACCACCGCCG 6559

Qy        334 SerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThr 353
Db       6560 AAACACCCGCCGACCTCGCCGGCACCTGGCTCGTCGTCGTGCCGGCCGCTCTGGACGACG 6619

Qy        354 TyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSer 373
Db       6620 ---ACGCGCTCCCCTCCGCGCTCA-----------------------CCGAGGCGCTCA 6652

Qy        374 AsnGlyLeuAlaArgSerGluGlu-------------------------SerAlaArg 384
Db       6653 CCCGGCGCGGCGCGCGCGTCCTCGCCTTGCGCCTGAGCCAGGCCCACCTGGACCGCGAGG 6712

Qy        385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys 404
Db       6713 CTC--------------TCGCCGAGCATCTGCGCCAGGCTTGCGCCGAGACCGCCCCGA 6757

Qy        405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424
```

```
Db   6758 TTCGCGGCGTGCTCTCGCTCCTCGCCCTCGACGAGCGCC------CCCTCGCAGACCGTC 6811
Qy    425 MetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrp 444
Db   6812 CTGCCCTGCCCGCCGGACTCGCCCTCTCGCTTTCTCTCGCTCAAGCCCTCGGCGACCTCG 6871
Qy    445 SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg 464
Db   6872 ACCTCGAGGCGCCCTTGTGGTTCTTCACGCGCGGCGCCGTCTCCATTGGACACTCTGACC 6931
Qy    465 GlyLysProPheProGluLysPro------------------------------------ 472
Db   6932 CCCTCGCCCATCCCGCCCAGGCCATGACCTGGGGCTTGGGCCGCGTCATCGGCCTCGAGC 6991
Qy    473 ---------ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer------ 487
Db   6992 ACCCCGACCGGTGGGGAGGTCTCGTCGACGTCTGCGCTGGGGTCGACGAGAGCGCCGTGG 7051
Qy    488 ------------------LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
Db   7052 GCCGCTTGCTGCCGGCCCTCGCCGAGCGCCACGACGAAGACCAGCTCGCTCTCCGCCCGG 7111
Qy    502 ProAspAsn---------SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu 518
Db   7112 CCGGACTCTACGCTCGCCGCATCGTCCGCGCCCCGCTCGGCGATGCGCCTCCCG------ 7165
Qy    519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIle 538
Db   7166 ------------------CGCGCGACTTCACGC--------------------------- 7180
Qy    539 LeuIleProSerProThrSerThrSerProSerValPro---ProValValSerLysGln 557
Db   7181 ------CCGGAGGCACCATTCTCATCACCGGCGGCACCGGCGCCATTGGCGCTCACGTCG 7234
Qy    558 Pro---GlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
Db   7235 CCCGATGGCTCGCTCGAAGAGGCGCTCAGCACCTCGTCCTCATCA--------------- 7279
Qy    577 GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596
Db   7279 ------------------------------------------------------------ 7279
Qy    597 PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla 616
Db   7280 ---------GCCGCCGAGGCGCCGAGGCCCCTG--------------------------- 7303
Qy    617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
Db   7304 -----------------GCGCCTCGGAGCTCCACGACGAGCTCTCGGCCCTCGGCGCGC 7345
Qy    637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln 656
Db   7346 GCACCACCCTCGCCGCGTGCGATGTCGCCGACCGGAATGCTGTCGCCACGCTTCTTGAGC 7405
Qy    657 LeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle------GlnLeu 674
Db   7406 AGCTCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGAGCGGCATCGAACACC 7465
Qy    675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis------------------ 688
Db   7466 ACGCTCCGCTCGACGCCACCTCTTTCAGGGATCTCGCCGAGGTTGTCTCCGGCAAGGTCG 7525
Qy    689 GlnIleArgHisHisLysThrAspLeuLeuSer------ProSerGlyThrGlyGluVal 706
Db   7526 AAGGTGCAAAG-CACCTCCACGACCTGCTCGGCTCTCGACCCCTCGACGCCTTTGTTCTC 7584
Qy    707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg--------- 723
Db   7585 TTTTCGTCCGGCGC-GGCCGTCTGGGGCGGCGGACAGCAAGGCGGCTACGCGGCCGCAAA 7643
Qy    724 ------------------------------------------------AspValLeu 726
Db   7644 CGCCTTCCTCGACGCCCTTGCCGAGCATCGGCGCAGCGCTGGATTGACAGCGACGTCGGT 7703
Qy    727 GlyLeuGlyAspVal-----------------------------------------Pro 732
Db   7704 GGCCTGGGGCGCGTGGGGCGGCGGCGGCATGGCCACCGATCAGGCGGCAGCCCACCTCCA 7763
Qy    733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet------------TrpValIl 748
Db   7764 ACAGCGCGGTCTGTCGCGGATGGCCCCCTCGCTTGCCCTGGCGGCGCTCGCGCTGGCTCT 7823
Qy    748 eAspValAspAspThr------------GlyAspGluAspTrpIleGluThrTyrProSe 764
Db   7824 GGAGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTC 7883
Qy    764 r---------------GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
Db   7884 GTTCAGCGCCGCTCGCCCCCGCCCGCTCCTGCGCGAT---TTGCCCGAGGCGCAGCGCGC 7940
Qy    779 rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLe 799
Db   7941 TCTCGAGACCAGCGAAGGCGCGTCCTCCGAGCATGGCCCGGCC----------------- 7983
Qy    799 uProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspAr 819
Db   7984 -CCCGACCTCCTCGACAAGCTCCGGAGCCGCTCGGAGAGCGAGCAGCTT---------CG 8033
Qy    819 gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
Db   8034 TCTGCTCGTCTCGCTGGTGCGCCACGAGACGGCCCTCGTCCTCGGCCACGAAGGCGCCTC 8093
Qy    839 ----------------------------------ValGluPheArgGlnTrp------- 844
Db   8094 CCATGTCGACCCCGACAAGGGCTTCCTCGATCTCGGTCTCGATTCGCTCATGGCCGTCGA 8153
Qy    844 ------------------------------------------------------------ 844
Db   8154 GCTTCGCCGGCGCTTGCAACAGGCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGA 8213
Qy    845 -----------ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV 861
Db   8214 CCATCCCTCTCCTCATCGAGTCGCGCTCTTCTTGCGCGACTCGC---------TCGCCCA 8264
Qy    861 alProPheLeuAlaGlyLeuProAspSer 870
Db   8265 CGCCCTCGGCACGAGGCTCTCCGTCGAGC 8293
```

RESULT 11

```
US-08-729-214-6
; Sequence 6, Application US/08729214
; Patent No. 5817502
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  van Pee, Karl-Heinz
;    APPLICANT:  Kirner, Sabine
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/729,214
;      FILING DATE:  TBA
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-729-214-6

Alignment Scores:
Pred. No.:              0.000564        Length:         28958
Score:                  178.50          Matches:        213
Percent Similarity:     29.87%          Conservative:   89
Best Local Similarity:  21.07%          Mismatches:     399
Query Match:            2.41%           Indels:         312
DB:                     1               Gaps:           42

US-09-913-878A-2 (1-1402) x US-08-729-214-6 (1-28958)

Qy        38 ProHisArgArgLys-----------------------------GluLeuAlaGluSer 47
Db      5663 CCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCC---TCG 5719

Qy        48 AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg 67
Db      5720 CCGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGA 5779

Qy        68 LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSer 87
Db      5780 TCGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCT 5839

Qy        88 Ser---------AsnTrpVal-------------------------ProLysAla--- 94
Db      5840 TCCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCA 5899

Qy        95 -----------HisAlaAspProAspThr-----------------LeuProTrpSer 104
Db      5900 CCTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACC 5959

Qy       105 LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeu 124
Db      5960 ACGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCG 6019

Qy       125 GluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArg 144
Db      6020 ACGGCGTACGTGCCCTTC---ACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTC 6076

Qy       145 ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp 164
Db      6077 ACGCTGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCAT 6136

Qy       165 GluProAla---------AsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
Db      6137 GCGCCTTCCTTCCCACCCTCCGCAAGGACGCGACGACGCCGAGGCGTTCACCGCCGCGC 6196

Qy       182 SerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProVal 201
Db      6197 TCGGCGCTCTCCACTCCG-----------------------CAGGCATCACACCCGACT 6232

Qy       202 AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
Db      6233 GGAGCGCTTTCTTCGCCCCCTTCGCTCCAC---------------------------- 6262

Qy       222 AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValPro---Ile 240
Db      6263 --------------------------GCAAGGTCTCCCTCCCCACCTATGCCTTCCAGC 6295

Qy       241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
Db      6296 GCGAGCGCTTCTGGCCCGACGCCTCCAAGGCACCCGGCG------------CCGACGTCA 6343

Qy       261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
Db      6344 GCCACCTTGCTCCGCTCGAGGGGGGGCTCTGGCAAGCCATCGAGCGCGGGGACCTCGATG 6403

Qy       281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
Db      6404 CGCTCAGCGGTCAGCTCCACGTGGACGGCGACGAGCGGCGCGCCGCGCTCGCCCTGCTCC 6463

Qy       301 AlaProProSer--------------------GlnProArgGluLysArgProValAsp 313
Db      6464 TTCCCACCCTCTCGAGCTTTCGCCACGAGCGGCAAGAGCAGAGCACGGTCGACGCCTGGC 6523

Qy       314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys 333
Db      6524 GCTACCGTA-----------------------TCACCTGGAAGCCTCTGACCACCGCCG 6559

Qy       334 SerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThr 353
Db      6560 AAACACCCGCCGACCTCGCCGGCACCTGGCTCGTCGTCGTGCCGGCCGCTCTGGACGACG 6619

Qy       354 TyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSer 373
Db      6620 ---ACGCGCTCCCCTCCGCGCTCA-----------------------CCGAGGCGCTCA 6652

Qy       374 AsnGlyLeuAlaArgSerGluGlu-------------------------SerAlaArg 384
Db      6653 CCCGGCGCGGCGCGCGTCCTCGCCTTGCGCCTGAGCCAGGCCCACCTGGACCGCGAGG 6712

Qy       385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys 404
Db      6713 CTC--------------TCGCCGAGCATCTGCGCCAGGCTTGCGCCGAGACCGCCCCGA 6757

Qy       405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424
Db      6758 TTCGCGGCGTGCTCTCGCTCCTCGCCCTCGACGAGCGCC------CCCTCGCAGACCGTC 6811

Qy       425 MetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrp 444
Db      6812 CTGCCCTGCCCGCCGGACTCGCCCTCTCGCTTTCTCTCGCTCAAGCCCTCGGCGACCTCG 6871

Qy       445 SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg 464
Db      6872 ACCTCGAGGCGCCCTTGTGGTTCTTCACGCGCGGCGCCGTCTCCATTGGACACTCTGACC 6931

Qy       465 GlyLysProPheProGluLysPro---------------------------------- 472
Db      6932 CCCTCGCCCATCCCGCCCAGGCCATGACCTGGGGCTTGGGCCGCGTCATCGGCCTCGAGC 6991

Qy       473 ---------ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer------ 487
Db      6992 ACCCCGACCGGTGGGGAGGTCTCGTCGACGTCTGCGCTGGGGTCGACGAGAGCGCCGTGG 7051

Qy       488 -----------------LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
Db      7052 GCCGCTTGCTGCCGGCCCTCGCCGAGCGCCACGACGAAGACCAGCTCGCTCTCCGCCCGG 7111

Qy       502 ProAspAsn---------SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu 518
Db      7112 CCGGACTCTACGCTCGCCGCATCGTCCGCGCCCCGCTCGGCGATGCGCCTCCCG------ 7165

Qy       519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIle 538
Db      7166 -----------------CGCGCGACTTCACGC-------------------------- 7180
```

```
Qy        539 LeuIleProSerProThrSerThrSerProSerValPro---ProValValSerLysGln 557
                  |||           :::||||||:::   |||   |||:::
Db       7181 ------CCGGAGGCACCATTCTCATCACCGGCGGCACCGGCGCCATTGGCGCTCACGTCG 7234

Qy        558 Pro---GlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
              |||   |||::::::||||||   :::         :::
Db       7235 CCCGATGGCTCGCTCGAAGAGGCGCTCAGCACCTCGTCCTCATCA--------------- 7279

Qy        577 GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596

Db       7279 ------------------------------------------------------------ 7279

Qy        597 PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla 616
                          ||||||   |||:::|||:::
Db       7280 ---------GCCGCCGAGGCGCCGAGGCCCCTG--------------------------- 7303

Qy        617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
                                |||         |||   |||      |||
Db       7304 -----------------GCGCCTCGGAGCTCCACGACGAGCTCTCGGCCCTCGGCGCGC 7345

Qy        637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln 656
                |||      |||                  :::               |||
Db       7346 GCACCACCCTCGCCGCGTGCGATGTCGCCGACCGGAATGCTGTCGCCACGCTTCTTGAGC 7405

Qy        657 LeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle------GlnLeu 674
                               :::                 :::|||
Db       7406 AGCTCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGAGCGGCATCGAACACC 7465

Qy        675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis------------------ 688
                |||   |||      :::      :::   |||
Db       7466 ACGCTCCGCTCGACGCCACCTCTTTCAGGGATCTCGCCGAGGTTGTCTCCGGCAAGGTCG 7525

Qy        689 GlnIleArgHisHisLysThrAspLeuLeuSer------ProSerGlyThrGlyGluVal 706
              :::::::::   |||      ||||||||         |||               :::
Db       7526 AAGGTGCAAAG-CACCTCCACGACCTGCTCGGCTCTCGACCCCTCGACGCCTTTGTTCTC 7584

Qy        707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg--------- 723
                :::   |||   |||||||:::   |||         |||:::|||:::|||
Db       7585 TTTTCGTCCGGCGC-GGCCGTCTGGGGCGGCGGACAGCAAGGCGGCTACGCGGCCGCAAA 7643

Qy        724 ---------------------------------------------------AspValLeu 726
                                                                 ||||||
Db       7644 CGCCTTCCTCGACGCCCTTGCCGAGCATCGGCGCAGCGCTGGATTGACAGCGACGTCGGT 7703

Qy        727 GlyLeuGlyAspVal------------------------------------------Pro 732
              |||||||||   |||                                          |||
Db       7704 GGCCTGGGGCGCGTGGGGCGGCGGCGGCATGGCCACCGATCAGGCGGCAGCCCACCTCCA 7763

Qy        733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet------------TrpValIl 748
              :::|||         |||         :::      :::                  ::
Db       7764 ACAGCGCGGTCTGTCGCGGATGGCCCCCTCGCTTGCCCTGGCGGCGCTCGCGCTGGCTCT 7823

Qy        748 eAspValAspAspThr------------GlyAspGluAspTrpIleGluThrTyrProSe 764
               ::::   |||:::|||               |||   ||||||            |||||
Db       7824 GGAGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTC 7883

Qy        764 r---------------GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
              |                  |||         |||   :::   :::   ||||||
Db       7884 GTTCAGCGCCGCTCGCCCCCGCCCGCTCCTGCGCGAT---TTGCCCGAGGCGCAGCGCGC 7940

Qy        779 rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLe 799
               ||||||               ||||||
Db       7941 TCTCGAGACCAGCGAAGGCGCGTCCTCCGAGCATGGCCCGGCC----------------- 7983

Qy        799 uProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspAr 819
               |||   |||   |||:::   |||   :::   :::      |||            ||
Db       7984 -CCCGACCTCCTCGACAAGCTCCGGAGCCGCTCGGAGAGCGAGCAGCTT---------CG 8033

Qy        819 gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
               ||||:::      |||   |||         :::            |||         |||
Db       8034 TCTGCTCGTCTCGCTGGTGCGCCACGAGACGGCCCTCGTCCTCGGCCACGAAGGCGCCTC 8093

Qy        839 ---------------------------------ValGluPheArgGlnTrp------- 844
                                               |||      |||   |||
Db       8094 CCATGTCGACCCCGACAAGGGCTTCCTCGATCTCGGTCTCGATTCGCTCATGGCCGTCGA 8153

Qy        844 ------------------------------------------------------------ 844

Db       8154 GCTTCGCCGGCGCTTGCAACAGGCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGA 8213

Qy        845 -----------ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV 861
                         :::   ||||||   ||||||   |||||||||
Db       8214 CCATCCCTCTCCTCATCGAGTCGCGCTCTTCTTGCGCGACTCGC---------TCGCCCA 8264

Qy        861 alProPheLeuAlaGlyLeuProAspSer 870
                |||         |||   |||   |||
Db       8265 CGCCCTCGGCACGAGGCTCTCCGTCGAGC 8293
```

```
RESULT 12
US-09-028-934-6
; Sequence 6, Application US/09028934
; Patent No. 6117670
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Lam, Steven T.
;    APPLICANT:  Hammer, Philip E.
;    APPLICANT:  van Pee, Karl-Heinz
;    APPLICANT:  Kirner, Sabine
;    APPLICANT:  Young, Thomas R.
;    TITLE OF INVENTION:  Pyrrolnitrin Biosynthesis Genes and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6117670artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/028,934
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1506/CIP7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-09-028-934-6

Alignment Scores:
Pred. No.:            0.000564      Length:       28958
```

```
Score:                    178.50     Matches:        213
Percent Similarity:       29.87%     Conservative:   89
Best Local Similarity:    21.07%     Mismatches:     399
Query Match:              2.41%      Indels:         312
DB:                       3          Gaps:           42

US-09-913-878A-2 (1-1402) x US-09-028-934-6 (1-28958)

Qy     38 ProHisArgArgLys------------------------------GluLeuAlaGluSer 47
Db   5663 CCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCC---TCG 5719
Qy     48 AspGluAspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArg 67
Db   5720 CCGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGA 5779
Qy     68 LysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSer 87
Db   5780 TCGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCT 5839
Qy     88 Ser---------AsnTrpVal--------------------------ProLysAla--- 94
Db   5840 TCCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCA 5899
Qy     95 ------------HisAlaAspProAspThr------------------LeuProTrpSer 104
Db   5900 CCTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACC 5959
Qy    105 LysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrValLeuLeu 124
Db   5960 ACGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCG 6019
Qy    125 GluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThrPheGlyArg 144
Db   6020 ACGGCGTACGTGCCCTTC---ACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTC 6076
Qy    145 ThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLysArgLysAsp 164
Db   6077 ACGCTGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCAT 6136
Qy    165 GluProAla---------AsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
Db   6137 GCGCCTTCCTTCCCACCCTCCGCAAGGGACGCGACGACGCCGAGGCGTTCACCGCCGCGC 6196
Qy    182 SerAlaThrGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProVal 201
Db   6197 TCGGCGCTCTCCACTCCG-----------------------CAGGCATCACACCCGACT 6232
Qy    202 AsnThrGlySerLysArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArg 221
Db   6233 GGAGCGCTTTCTTCGCCCCCTTCGCTCCAC---------------------------- 6262
Qy    222 AlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProValPro---Ile 240
Db   6263 ---------------------------GCAAGGTCTCCCTCCCCACCTATGCCTTCCAGC 6295
Qy    241 AlaSerAlaLeuAspLysValProThrArgArgHisAlaAsnThrArgAspProThrAla 260
Db   6296 GCGAGCGCTTCTGGCCCGACGCCTCCAAGGCACCCGGCG------------CCGACGTCA 6343
Qy    261 ThrGlyHisArgArgAlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyr 280
Db   6344 GCCACCTTGCTCCGCTCGAGGGGGGGCTCTGGCAAGCCATCGAGCGCGGGGACCTCGATG 6403
Qy    281 GlySerSerValPheSerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGlu 300
Db   6404 CGCTCAGCGGTCAGCTCCACGTGGACGGCGACGAGCGGCGCGCCGCGCTCGCCCTGCTCC 6463
Qy    301 AlaProProSer--------------------GlnProArgGluLysArgProValAsp 313
Db   6464 TTCCCACCCTCTCGAGCTTTCGCCACGAGCGGCAAGAGCAGAGCACGGTCGACGCCTGGC 6523
Qy    314 AlaThrValPheGluAlaGlyHisLeuIleGluSerProSerLysGlyArgThrThrLys 333
Db   6524 GCTACCGTA-----------------------TCACCTGGAAGCCTCTGACCACCGCCG 6559
Qy    334 SerHisIleAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThr 353
Db   6560 AAACACCCGCCGACCTCGCCGGCACCTGGCTCGTCGTCGTGCCGGCCGCTCTGGACGACG 6619
Qy    354 TyrTyrGluSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSer 373
Db   6620 ---ACGCGCTCCCCTCCGCGCTCA-----------------------CCGAGGCGCTCA 6652
Qy    374 AsnGlyLeuAlaArgSerGluGlu--------------------------SerAlaArg 384
Db   6653 CCCGGCGCGGCGCGCGTCCTCGCCTTGCGCCTGAGCCAGGCCCACCTGGACCGCGAGG 6712
Qy    385 SerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLys 404
Db   6713 CTC---------------TCGCCGAGCATCTGCGCCAGGCTTGCGCCGAGACCGCCCCGA 6757
Qy    405 PheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPhe 424
Db   6758 TTCGCGGCGTGCTCTCGCTCCTCGCCCTCGACGAGCGCC------CCCTCGCAGACCGTC 6811
Qy    425 MetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrp 444
Db   6812 CTGCCCTGCCCGCCGGACTCGCCCTCTCGCTTTCTCTCGCTCAAGCCCTCGGCGACCTCG 6871
Qy    445 SerThrAlaArgAspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArg 464
Db   6872 ACCTCGAGGCGCCCTTGTGGTTCTTCACGCGCGGCGCCGTCTCCATTGGACACTCTGACC 6931
Qy    465 GlyLysProPheProGluLysPro------------------------------------ 472
Db   6932 CCCTCGCCCATCCCGCCCAGGCCATGACCTGGGGCTTGGGCCGCGTCATCGGCCTCGAGC 6991
Qy    473 ---------ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSer------ 487
Db   6992 ACCCCGACCGGTGGGGAGGTCTCGTCGACGTCTGCGCTGGGGTCGACGAGAGCGCCGTGG 7051
Qy    488 -----------------LysGlySerAlaValValLeuSerAlaValLeuAspTyrAsn 501
Db   7052 GCCGCTTGCTGCCGGCCCTCGCCGAGCGCCACGACGAAGACCAGCTCGCTCTCCGCCCGG 7111
Qy    502 ProAspAsn---------SerProThrAlaProLeuTyrLeuValLysLeuLysProLeu 518
Db   7112 CCGGACTCTACGCTCGCCGCATCGTCCGCGCCCCGCTCGGCGATGCGCCTCCCG------ 7165
Qy    519 MetPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIle 538
Db   7166 -----------------CGCGCGACTTCACGC-------------------------- 7180
Qy    539 LeuIleProSerProThrSerThrSerProSerValPro---ProValValSerLysGln 557
Db   7181 ------CCGGAGGCACCATTCTCATCACCGGCGGCACCGGCGCCATTGGCGCTCACGTCG 7234
Qy    558 Pro---GlyAlaValGluGluValIleGlnTrpLeuThrMetGlyGlnHisSerLeuVal 576
Db   7235 CCCGATGGCTCGCTCGAAGAGGCGCTCAGCACCTCGTCCTCATCA-------------- 7279
Qy    577 GlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuArgGlu 596
Db   7279 ------------------------------------------------------------ 7279
Qy    597 PheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePheAla 616
Db   7280 ---------GCCGCCGAGGCGCCGAGGCCCCTG------------------------- 7303
Qy    617 GluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAlaGlu 636
Db   7304 -----------------GCGCCTCGGAGCTCCACGACGAGCTCTCGGCCCTCGGCGCGC 7345
Qy    637 GluProValGluGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGln 656
```

```
Db       7346 GCACCACCCTCGCCGCGTGCGATGTCGCCGACCGGAATGCTGTCGCCACGCTTCTTGAGC 7405

Qy        657 LeuAspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIle------GlnLeu 674
                                 :::                :::|||
Db       7406 AGCTCGACGCCGAAGGGTCGCAGGTCCGCGCCGTGTTCCACGCGAGCGGCATCGAACACC 7465

Qy        675 GlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHis------------------ 688
                 |||    |||      :::     :::  |||
Db       7466 ACGCTCCGCTCGACGCCACCTCTTTCAGGGATCTCGCCGAGGTTGTCTCCGGCAAGGTCG 7525

Qy        689 GlnIleArgHisHisLysThrAspLeuLeuSer------ProSerGlyThrGlyGluVal 706
              :::::::::    |||       ||||||||         |||                :::
Db       7526 AAGGTGCAAAG-CACCTCCACGACCTGCTCGGCTCTCGACCCCTCGACGCCTTTGTTCTC 7584

Qy        707 MetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg--------- 723
                 :::   |||    ||||||:::    |||      |||:::|||:::|||
Db       7585 TTTTCGTCCGGCGC-GGCCGTCTGGGGCGGCGGACAGCAAGGCGGCTACGCGGCCGCAAA 7643

Qy        724 ---------------------------------------------------AspValLeu 726
                                                                 ||||||
Db       7644 CGCCTTCCTCGACGCCCTTGCCGAGCATCGGCGCAGCGCTGGATTGACAGCGACGTCGGT 7703

Qy        727 GlyLeuGlyAspVal-----------------------------------------Pro 732
              |||||||||   |||                                         |||
Db       7704 GGCCTGGGGCGCGTGGGGCGGCGGCGGCATGGCCACCGATCAGGCGGCAGCCCACCTCCA 7763

Qy        733 SerAlaValGln-GlyArgPheGlySerAlaLysGlyMet-----------TrpValIl 748
              :::|||       |||          :::      :::                  ::
Db       7764 ACAGCGCGGTCTGTCGCGGATGGCCCCCTCGCTTGCCCTGGCGGCGCTCGCGCTGGCTCT 7823

Qy        748 eAspValAspAspThr-----------GlyAspGluAspTrpIleGluThrTyrProSe 764
               ::::   |||:::|||             |||    ||||||                |||||
Db       7824 GGAGCACGACGAGACCACCGTCACCGTCGCCGACATCGACTGGGCGCGCTTTGCGCCTTC 7883

Qy        764 r---------------GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgTh 779
              |                  |||             |||    :::   :::   ||||||
Db       7884 GTTCAGCGCCGCTCGCCCCCGCCCGCTCCTGCGCGAT---TTGCCCGAGGCGCAGCGCGC 7940

Qy        779 rLeuGluValArgSerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLe 799
               ||||||             ||||||
Db       7941 TCTCGAGACCAGCGAAGGCGCGTCCTCCGAGCATGGCCCGGCC----------------- 7983

Qy        799 uProValLeuGluAspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspAr 819
               |||   |||   |||:::   |||   :::   :::      |||            ||
Db       7984 -CCCGACCTCCTCGACAAGCTCCGGAGCCGCTCGGAGAGCGAGCAGCTT---------CG 8033

Qy        819 gLeuIleAsnAspLeuGlnArgGlnPheSer-GluGlnLysHisAlaLeuAsnArgPro- 838
              ||||:::      |||   |||       :::            |||            |||
Db       8034 TCTGCTCGTCTCGCTGGTGCGCCACGAGACGGCCCTCGTCCTCGGCCACGAAGGCGCCTC 8093

Qy        839 -----------------------------------ValGluPheArgGlnTrp------- 844
                                                 |||    |||  |||
Db       8094 CCATGTCGACCCCGACAAGGGCTTCCTCGATCTCGGTCTCGATTCGCTCATGGCCGTCGA 8153

Qy        844 ------------------------------------------------------------ 844

Db       8154 GCTTCGCCGGCGCTTGCAACAGGCCACCGGCATCAAGCTCCCGGCCACCCTCGCCTTCGA 8213

Qy        845 -----------ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgV 861
                         :::   ||||||    ||||||    |||||||||
Db       8214 CCATCCCTCTCCTCATCGAGTCGCGCTCTTCTTGCGCGACTCGC---------TCGCCCA 8264

Qy        861 alProPheLeuAlaGlyLeuProAspSer 870
                 |||      |||   |||   |||
Db       8265 CGCCCTCGGCACGAGGCTCTCCGTCGAGC 8293
```

RESULT 13
US-09-220-132-30
; Sequence 30, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITONS FOR THE IDENTIFICATION AND ASSESSMENT
; TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CAN(
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30
;   LENGTH: 14770
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-220-132-30

Alignment Scores:			
Pred. No.:	0.000461	Length:	14770
Score:	174.00	Matches:	277
Percent Similarity:	30.75%	Conservative:	208
Best Local Similarity:	17.56%	Mismatches:	551
Query Match:	2.35%	Indels:	541
DB:	4	Gaps:	69

US-09-913-878A-2 (1-1402) x US-09-220-132-30 (1-14770)

```
Qy         30 CysValAlaAspThrThrLeuThrProHisArgArgLysGluLeuAlaGluSerAspGlu 49
              |||      :::   ||||||      |||:::::::::         ::::::   |||:::
Db       4480 TGCAACTTAAATATCACTCTGCCAGCACATAAAAAGGAGACAGAGTCAGATCAAGATGAT 4539

Qy         50 AspPheGlyArgHisAspLysIleTyrArgAlaLeuAsnPheLeuTyrTrpArgLysAsp 69
              :::      :::   |||:::               :::          ::::::
Db       4540 GAGATTGAGAAAACAGATAGACGACAGAGCTTCGCATCCTTAGCTTTACGTAAGCGCTAC 4599

Qy         70 AspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAlaAlaSerSerAsn 89
                 |||   :::               ||||||               ::::::
Db       4600 AGCTACTTGACTGAGCCTGGA---------ATGATTGAACGGAGTACAGGAGCAACAAGA 4650

Qy         90 TrpValProLysAlaHisAla--------------------------------------- 96
                 :::|||    ::::::
Db       4651 TCCCTCCCCACCACTTACTCATACAAGCCATTCTTTTCTACAAGACCATACCAGTCCTGG 4710

Qy         97 ------AspProAspThrLeuProTrp--------------------------------- 103
                       |||   |||:::|||
Db       4711 ACAACAGCTCCGATTACAGTGCCTGGGCCAGCCAAGTCAGGCTTCACTTCCTTATCAAGT 4770

Qy        104 ---SerLysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThrVal 122
                 |||   |||      |||   |||:::               |||::::::   |||
Db       4771 TCTTCCTCTAATACGCCATCAGCTTCTCCGTTAAAATCAATATGGTCTGTTTCGACA--- 4827

Qy        123 LeuLeuGluValLeuAsnArgPheMetProProPro---AsnAsnThrProGlyArgThr 141
                                         |||   |||      :::|||   |||   :::
Db       4828 --------------------------CCTTCTCCAATCAAATCCACATTAGGCGCGTCA 4860

Qy        142 PheGlyArgThrLeuSerGlyProSerGlyLeuSerArgProThrSerThrAsnThrLys 161
                    ::::::         |||   ::::::   |||          :::   :::
Db       4861 ACTACATCTTCAGTTAAATCCATTAGTGACGTGGCATCTCCAATT---AGATCCTTACGG 4917

Qy        162 ArgLysAspGluProAlaAsnValThrPheAlaAspProProLysArgSerLeuThrArg 181
                          |||             :::       |||
Db       4918 ACAATGTCTTCGCCGATAAAAACTGTGGTGTCACAATCTCCATACAATATCCAGGTTTCC 4977

Qy        182 SerAlaThr---------------------------GlyProProIleHisGlyAlaAla 192
              |||   |||                              |||:::   |||   |||
Db       4978 TCTGGTACCCTGGCTAGAGCTCCAGCAGTCACGGAAGCTACGCCCTTAAAAGGGCTGGCA 5037

Qy        193 IleProLeuLysPhe---------ProAspProValAsnThrGlySerLys--------- 206
                       |||                  ||||||   |||
Db       5038 TCCAATTCTACGTTTTCCTCTCGAACCTCTCCAGTGACTACAGCAGGGTCTCTTTTGGAG 5097

Qy        207 --------------------ArgProSerLeuGluSerGluAsnLeuAsnGlnCysThr 219
                                  |||:::         |||:::|||            :::
Db       5098 AGGTCATCAATTACTATGACACCCCCTGCCTCCCCCAAATCAAACATTAATATGTATTCC 5157
```

```
Qy      220 LysArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAlaAlaAlaProProVal--- 238
                           |||          ::::::      :::||||||||    :::
Db     5158 TCAAGTTTGCCATTTAAG---------TCAATTATTACATCAGCAGCACCGCTAATATCT 5208

Qy      239 -----------------------ProIleAlaSerAlaLeuAsp-------------- 245
                                   |||:::   |||   :::|||
Db     5209 TCACCTTTAAAGTCAGTGGTGTCTCCAGTTAAATCACGAGTTGATGTCATTTCATCAGCC 5268

Qy      246 LysValProThrArgArgHisAlaAsnThrArgAspProThrAlaThrGlyHisArgArg 265
            |||:::                   ::::::                  ||||||
Db     5269 AAAATTACAATGGCATCTTCTCTCTCATCACCTGTGAAGCAGATGCCTGGACATGCAGAG 5328

Qy      266 AlaAspGlnValAspSerPheAspThrSerGlnGlyThrSerTyrGlySerSerValPhe 285
                     |||:::         :::         :::      |||::::::
Db     5329 GTAGCATTAGTCAATGGATCTATTTCCCCTCTAAAATATGCATCATCCTCAACTTTAATT 5388

Qy      286 SerAlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGluAlaProProSerGln 305
            :::      |||:::         :::      :::               |||      |||
Db     5389 AATGGATGCAAAGCCACTGCCACGTTACAGGAAAAAATTTCTTCTGCTACAAACTCT--- 5445

Qy      306 ProArgGluLysArgProValAspAlaThrValPheGluAlaGlyHisLeuIleGluSer 325
                              |||   :::   |||      |||         :::|||
Db     5446 -----------------GTGAGCTCTGTGGTCAGTGCAGCCACTGACACAGTTGAGAAA 5487

Qy      326 ProSerLysGlyArgThrThr------------------LysSerHisIleAspAsnGln 339
                           |||                     :::|||::::::
Db     5488 GTGTTTTCTACCACGACTGCAATGCCATTTTCCCCACTCAGGTCATATGTTTCTGCAGCA 5547

Qy      340 ProLeuSerSerSerSerGlnGlyGluThrSerPheSerThrTyrTyrGluSerPhePro 359
            |||   :::         |||               |||   |||      |||   |||
Db     5548 CCATCAGCTTTTCAGTCTCTAAGAACTCCTTCCGCAAGTGCACTCTATACATCCCTTGGG 5607

Qy      360 SerSer---------------------------------------------------- 361
            ||||||
Db     5608 TCGTCAATATCTGCAACTACCTCATCTGTAACTTCATCAATTATAACAGTGCCAGTATAC 5667

Qy      362 GlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLeuAlaArgSerGluGlu 381
                           :::|||||||||:::                  |||      |||
Db     5668 TCTGTAGTCAATGTTTTGCCAGAACCAGCATTAAAGAAA---CTTCCAGACTCTAATTCA 5724

Qy      382 SerAlaArgSerGlnValGlnValHisAlaProValValAlaAlaArgLeuArgAsnIle 401
                  :::|||         :::   :::|||:::
Db     5725 TTTACAAAATCAGCAGCAGCCTTGCTGTCACCCATTAAAACATTGACTACGGAGACA--- 5781

Qy      402 TrpProLysPheProLysTrpLeuHisGluAlaProLeuAlaValAlaTrpGluValThr 421
                                     |||         |||            :::   |||
Db     5782 -----------------------CATCCTCAGCCT---------CACTTCAGTCGAACT 5808

Qy      422 ArgLeuPheMetHisCysLysValAspLeuGluAspGluSerLeuGlyLeuLysTyrAsp 441
                  :::   :::   |||         :::|||   |||
Db     5809 TCATCTCCAGTTAAGTCATCTTTGTTCCTTGCACCCTCTGCCCTTAAGTTGTCTACACCA 5868

Qy      442 ProSerTrpSerThrAla------------ArgAspValThrAspIleTrpLysThrLeu 457
               |||   |||::::::            :::||||||   ::::::   :::   |||
Db     5869 TCTTCTTTATCTTCCAGTCAGGAGATACTAAAAGATGTAGCTGAAATGAAAGAGGACCTA 5928

Qy      458 TyrArgLeuAspAlaPhe-----------------ArgGlyLysProPheProGluLys 471
               |||:::   |||                    ||||||||             :::
Db     5929 ATGCGGATGACCGCAATACTACAGACAGATGTGCCTGAGGAGAAGCCATTCCAACCTGAA 5988

Qy      472 ProProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSerLysGlySerAla 491
               |||   :::                  |||      :::   :::
Db     5989 CTCCCAAAGGAA--------------------GGGAGAATAGATGATGAAGAACCTTTC 6027

Qy      492 ValValLeuSerAlaValLeuAspTyrAsnProAspAsnSerProThrAlaProLeuTyr 511
               ::::::   |||   :::
Db     6028 AAAATTGTAGAGAAAGTAAAGGAA-------------------------------GAC 6054

Qy      512 LeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrArg--------- 528
            |||||||||:::         ::::::   :::   |||         :::
Db     6055 TTAGTGAAAGTTAGTGAAATCCTTAAAAAGGATGTATGTGTAGATAATAAAGGATCACCC 6114

Qy      529 ----------------------------------------------ArgPheGlyPro 532
                                                              |||
Db     6115 AAATCACCAAAGAGTGACAAAGGACACTCTCCTGAAGATGACTGGATAGAATTTAGTTCG 6174

Qy      533 AspArgPhePheGluIleLeuIleProSerProThrSerThrSerProSerValProPro 552
            :::         |||            :::         |||   |||||||||:::|||
Db     6175 GAAGAAATCCGGGAAGCCAGACAACAAGCTGCTGCGAGCCAGTCTCCATCTCTGCCAGAG 6234

Qy      553 ValValSerLysGlnProGlyAlaValGluGlu--------------------ValIle 565
               |||      :::      |||      |||                     ||||||
Db     6235 AGAGTGCAAGTAAAAGCAAAAGCCGCCTCCGAAAAGGATTATAACTTGACCAAAGTTATT 6294

Qy      566 GlnTrpLeuThr------MetGlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPhe 583
               :::||||||         :::|||         ||||||            :::::::::
Db     6295 GATTACCTAACAAATGATATTGGGAGTAGTTCACTGACAAACTTAAAATACAAGTTTGAG 6354

Qy      584 PheAlaLysAspAlaGly-----------------TyrArgLysPro---------Leu 594
               ||||||   |||                        ||||||            |||
Db     6355 GATGCAAAGAAGGATGGTGAGGGAGGACAGAAAAGAGTTTAAAACCAGCAATTGCTTTG 6414

Qy      595 ArgGluPheGlnLeuArg-----------------------AlaGluAspProLysPro 606
            :::|||   :::|||:::                           :::      |||
Db     6415 CAGGAACACAAACTCAAAATGCCTCCAGCCTCCATGAGGACTTCCACCTCTGAGAAAGAA 6474

Qy      607 IleIleLysGluArgValHisPhePheAlaGluThrGlyIleThrPheArgProAspVal 626
            :::   |||         ||||||               |||         ||||||
Db     6475 TTGTGTAAAATGGCTGATTCCTTTTTTGGAACAGATACTATTTTAGAGTCTCCTGATGAC 6534

Qy      627 PheLysThrArgSerValValProAlaGluGluProValGluGlnArg-ThrGluPheLy 646
            |||                                          |||   |||:::   ||
Db     6535 TTT----------------------------------TCTCAACACGACCAAGATAAA 6558

Qy      646 sValSerGlnMetLeuAspTrpLeuLeuGlnLeuAspAsnAsnThrTrpGlnProHisLe 666
            ||||   :::         |||   |||                  |||||||||
Db     6559 AGTCCCTTGTCTGACAGTGGCTTTGAAACAAGAAGTGAAAAGACACCTTCAGCCCCAC-- 6616

Qy      666 uLysLeuPheSerArgIleGlnLeuGly------LeuSerLysThrTyrAlaIleMetTh 684
             |||         |||:::   |||                        :::::::::
Db     6617 -AAAGCGCTGAAACGACTGGTCCTAAACCACTTTTCATGAAGTTCCCATCCCTCCTGTT 6675

Qy      684 rLeuGluProHisGlnIleArgHis--HisLysThrAspLeuLeuSerProSerGlyThr 703
             |||:::      :::::::::   |||                        ||||||
Db     6676 ATTACAGAAACAAGAACTGAAGTGGTTCATGTTATCAGGAGCTATGATCCCTCAGCT--- 6732

Qy      704 GlyGluValMetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArg 723

Db     6732 ------------------------------------------------------------ 6732

Qy      724 AspValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGlySerAlaLys 743
                              ||||||||||||
Db     6733 --------------GGGGATGTTCCC------------------------------- 6744

Qy      744 GlyMetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrPro 763
                                        |||      |||:::   :::            |||
Db     6745 -----------------------CAGACCCAACCAGAGGAGCCTGTGTCACCTAAACCT 6780

Qy      764 SerGlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArg 783
            |||      |||                              ||||||   :::
Db     6781 TCACCTACTTTTATGGAA-----------------------------TTGGAACCAAAG 6810

Qy      784 SerValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGlu 803
                  |||   :::|||
Db     6811 CCCACCACCTCTAGTATTAAA--------------------------------------- 6831

Qy      804 AspArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAsp 823
                     :::|||||||||   |||
Db     6832 ------------GAAAAGGTTAAAGCATTTCAA------------------------- 6852

Qy      824 LeuGlnArgGlnPheSerGluGlnLysHis------AlaLeuAsnArgProValGluPhe 841
            ::::::            |||      |||         |||::::::   :::
Db     6853 ATGAAAGCCAGTAGTGAAGAAGATGACCACAATCGGGTTTTAAGCAAAGGCATGCGTGTT 6912

Qy      842 ArgGlnTrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgVal 861
```

```
                 ::::::          |||::::::  :::     ||||||:::   :::
Db      6913 AAAGAA---------GAGACTCACATAACCACAACCACCAGAATGGTTTATCATTCTCCA 6963

Qy       862 ProPheLeuAlaGlyLeuProAspSerGlnGluGluThrLeuAsnPhe----------- 877
             |||        |||                 ||||||||::::::
Db      6964 CCAGGCGGTGAAGGTGCATCTGAAAGAATTGAAGAAACCATGTCAGTCCATGACATCATG 7023

Qy       878 ---LeuMetAsnSerGlyPheAspProLysLysGln---------------LysTyrLeu 891
                         ||||||  ||||||  |||:::               ::::::
Db      7024 AAGGCCTTTCAGTCCGGGCGGGATCCTTCCAAAGAACTGGCAGGTCTGTTTGAACATAAG 7083

Qy       892 GlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThr------------------ 905
                :::::: |||:::   |||         :::|||
Db      7084 TCGGCAGTGTCTCCAGATGTTCACAAGTCTGCTGAAACCTCAGCCCAGCATGCAGAG 7143

Qy       906 ------------LeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyr 921
                         :::|||  ||||||          :::
Db      7144 AAGGACAACCAAATGAAACCCAAACTGGAGCGTATAATA--------------------- 7182

Qy       922 MetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSer 941
                                                 |||||||||:::
Db      7183 ----------------------------------GAAGTCCACATC------------ 7194

Qy       942 LysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAla 961
                                 |||                         :::|||::::::
Db      7195 ------------------GAAAAAGGTAACCAAGCTGAGCCCACTGAAGTCATTATTAGA 7236

Qy       962 ArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGlu 981
             :::      |||                                       ||||||
Db      7237 GAAACCAAAAAGCAT---------------------------------------CCAGAA 7257

Qy       982 LeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLys 1001
                         |||:::::::  ::::::                          |||
Db      7258 ------------AAAGAAATGTATGTATATCAG------------------------AAA 7281

Qy      1002 LysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIle 1021
                ||||||  ||||||
Db      7282 GACTTATCCGGGGAGATATT---------------------------------------- 7302

Qy      1022 ValAspGlyPheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLys 1041
                                                      :::|||      :::|||
Db      7303 ---------------------------------------AACCTAAAAGATTTTCTGCCA 7323

Qy      1042 LysAspLysThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLys 1061
             :::            |||             |||        |||           :::
Db      7324 GAAAAACACGATGCTTTTCCT---------TGTTCAGAGGAACAGGGTCAGCAAGAAGAA 7374

Qy      1062 GluGlnThrThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPhe 1081
             |||:::     |||    :::
Db      7375 GAAGAACTTACTGCTGAA------------------------------------------ 7392

Qy      1082 LeuGlyMetCysThrAsnTyrLysGluArgLeuCysTyrIleAsnAsnSerValSerAsn 1101
                                                                      :::
Db      7393 ------------------------------------------------------GAGTCA 7398

Qy      1102 LysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSerLysGlnGly 1121
              |||:::   :::    ||||||  |||         |||    |||    :::
Db      7399 TTGCCTTCTTATCTGGAGTCTTCCAGAGTAAACACTCCTGTGTCCCAAGAAGAAGAT--- 7455

Qy      1122 IleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeu 1141
                         :::        |||   |||||||||
Db      7456 ---------AGCCGCCCTAGTTCTGCTCAACTC-------------------------- 7479

Qy      1142 SerLeuProAspProMetTyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHis 1161
                :::   |||      ||||||:::         |||   :::              |||
Db      7480 ---ATATCTGATGACTCTTATAAAACATTGAAGCTTTTGAGTCAA-----------CAC 7524

Qy      1162 IleIleAspTyr------------------------------------LeuLysPheSer 1169
                |||:::|||                                   :::|||:::
Db      7525 TCAATAGAATACCATGACGATGAGTTGTCAGAACTAAGAGGGGAGTCTTACAGGTTTGCT 7584

Qy      1170 IleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAla 1189
                   :::      :::    ::::::|||:::         |||:::
Db      7585 ---GAGAAAATGCTTCTGTCAGAAAAGCTAGATGTGTCTCATTCT-------------- 7626

Qy      1190 LysAspThrGluAspGlyAla------HisPheTrpAspProAspLeuAlaSerTyrTyr 1207
                |||||||||:::             |||       |||
Db      7627 ---GATACTGAGGAATCGGTTACAGACCATGCAGGACCCCCTAGC--------------- 7668

Qy      1208 ThrPhePheLysGluIle------SerAspLysSerArgSerSerAlaLeuLeuPheThr 1225
                      |||:::          |||||||||  ||||||           :::   |||
Db      7669 ---------TCAGAGTTACAGGGGTCTGATAAG---CGGTCCAGAGAAAAAATAGCCACT 7716

Qy      1226 ThrLeuLysAsnArgIleGlyGluValGluLysGluTyrGlyArgLeuValLysAsnLys 1245
                |||      |||       :::   |||   |||        :::    :::|||
Db      7717 GCCCCCAAAAAAGAAATT------CTCTCCAAAATCTATAAAGATGTTTCTGAAAATGGT 7770

Qy      1246 GluMetArgAspSerLysAspProTyrProValArgValAsnGlnVal------------ 1261
                :::   |||||||||    :::       :::|||      :::
Db      7771 GTAGGTAAAGTGTCTAAAGATGAGCATTTTGATAAAGTGACAGTGTTGCACTATTCTGGC 7830

Qy      1262 ------------------TyrGluLysTrpCysAlaIleThrProGluAlaMetAspLys 1275
                            :::      |||        |||     :::    :::|||:::
Db      7831 AATGTTAGTAGTCCAAAACATGCCATGTGGATGCGCTTTACTGAGGACAGATTAGACAGA 7890

Qy      1276 SerGlyAlaAsn------TyrAspSerLysValIleArg--------------------- 1286
                               |||:::   :::|||   |||
Db      7891 GGTAGAGAGAAGTTGATATATGAAGATAGGGTGGACAGGACTGTGAAGGAGGCTGAAGAA 7950

Qy      1287 ---LeuLeuGluLeuSer---PheLeuAlaAspArg------------------GluMet 1298
                |||   |||:::|||   |||      |||:::                   |||:::
Db      7951 AAACTGACTGAAGTGTCACAGTTTTTCGTGACAAAACTGAAAAGCTAAATGATGAACTG 8010

Qy      1299 AsnThrTrpAlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerPro 1318
             :::                                 |||   |||       :::||||||
Db      8011 CAGTCCCCAGAGAAAAAGGCACGCCCTAAAAATGGCAAAGAATATTCTTCTCAAAGCCCT 8070

Qy      1319 Lys--------------------PheValTrpGlnMetAlaGlyArgGlnLeuAlaTyr 1331
                                    :::   ::::::                    :::
Db      8071 ACCAGTAGCAGCCCTGAGAAAGTGCTACTGACAGAACTGCTGGCATCCAATGATGAGTGG 8130

Qy      1332 IleLysAlaGlnMetThrSerArgProGlyGluGly--------------------Ala 1344
             :::||||||:::                |||:::|||                   |||
Db      8131 GTTAAGGCAAGACAGCATGGCCCTGATGGACAAGGCTTCCCCAAGGCCGAGGAGAAGGCA 8190

Qy      1345 ProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLysLysPheThrLys 1364
             |||:::|||   ::::::                     |||:::|||
Db      8191 CCCAGTCTGCCCAGCAGC---------------------CCAGAGAAGATGGTTCTCTCC 8229

Qy      1365 GlnTyrVal------------AlaArgLeuGluGlyAspGlySer 1375
             |||                  :::   :::|||   ||||||
Db      8230 CAACAGACTGAGGACAGCAAGTCCACAGTGGAAGCCAAAGGAAGT 8274
```

RESULT 14
US-09-426-436-1/c
; Sequence 1, Application US/09426436
; Patent No. 6225066
; GENERAL INFORMATION:
;   APPLICANT: William R. Jacobs, Jr.
;   APPLICANT: Barry R. Bloom
;   APPLICANT: Graham F. Hatfull
;   TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
;   TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Amster, Rothstein & Ebenstein
;   STREET: 90 Park Avenue
;   CITY: New York
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 10016
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
;   COMPUTER: IBM PC Compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Word Processor (ASCII)

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/426,436
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/705,557
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/057,531
;      FILING DATE:
;      APPLICATION NUMBER:  07/833,431
;      FILING DATE:  February 7, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pasqualini, Patricia A.
;      REGISTRATION NUMBER:  34,894
;      REFERENCE/DOCKET NUMBER:  96700/238
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  52297
;      TYPE:  nucleotide
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  phage genome sequence
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    FRAGMENT TYPE:  not applicable.
;    ORIGINAL SOURCE:
;      ORGANISM:  mycobacteriophage L5
;      STRAIN:  not applicable
;      INDIVIDUAL ISOLATE:  L5
;      DEVELOPMENTAL STAGE:  not applicable
;      HAPLOTYPE:  not applicable
;      TISSUE TYPE:  not applicable
;      CELL TYPE:  not applicable
;      CELL LINE:  not applicable
;      ORGANELLE:  not applicable
;    IMMEDIATE SOURCE:  mycobacteriophage L5 particles
;    POSITION IN GENOME:  entire genome
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    PUBLICATION INFORMATION:
;      AUTHORS:  Hatfull and Sarkis
;      TITLE:  DNA Sequence, Structure and Gene
;      TITLE:  Expression of Mycobacteriophage L5:
;      TITLE:  A Phage System for Mycobacterial
;      TITLE:  Genetics
;      JOURNAL:  Molecular Microbiology
;      VOLUME:  7
;      PAGES:  395-405
;      DATE:  1993
US-09-426-436-1

Alignment Scores:
Pred. No.:              0.00719       Length:        52297
Score:                  171.00        Matches:       308
Percent Similarity:     32.70%        Conservative:  192
Best Local Similarity:  20.14%        Mismatches:    524
Query Match:            2.31%         Indels:        513
DB:                     3             Gaps:          76

US-09-913-878A-2 (1-1402) x US-09-426-436-1 (1-52297)

Qy         35 ThrLeuThrProHisArgArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHis 54
Db      44666 ACGCGGAAACCTCACGATAAGTCGAAGGCTTTTCAGAGGCTCCGGGATCACGGCGAGTAC 44607

Qy         55 --------------------------------AspLysIleTyrArgAlaLeuAsnPhe 63
Db      44606 ATGCGAGTGGGTAGCTTCCGCACGATGTCCGAAGACAAGAAGAAGCGCCTATTGTCGTGG 44547

Qy         64 LeuTyrTrpArgLysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGlu--- 82
Db      44546 TGGAAGATGCTCCGCGACGACGACTTA---GTGCTCGAGTTCGACCCATCCATCGAGCCC 44490

Qy         83 ---------AlaLysAlaAlaSerSerAsnTrpValProLysAlaHisAlaAspProAsp 99
Db      44489 TACGAAGGTATGGCCGGCGGCGGGTTCAGGTACGTCCCCCGAGGCATCGAGGATGACGAT 44430

Qy        100 ThrLeu------------ProTrpSerLys-GluProProArgAlaAlaThrAlaGlyGl 115
Db      44429 CTCCTGATCCGGGTGAACGAGCACACCAACCTCACCGCCGAGGGTGAACTCCTCTGGTCG 44370

Qy        115 nGlnTrpAlaLeuGlnThrValLeuLeuGluValLeuAsnArgPheMetProProProAs 135
Db      44369 T---------------------------------------------------------GG 44367

Qy        135 nAsnThrProGlyArgThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArgPr 155
Db      44366 CCAGACGACATCGAGGAGCTTCTCTCGGAGCCCTAAAAGTCACGACCGGTTGTGTGAGCC 44307

Qy        155 oThrSerThrAsnThrLysArgLysAspGluProAlaAsnValThr-------------- 170
Db      44306 AACCCAGGCC---GCGCAACGAAAGGAAACCCCGTGCCC-CTCACTCGGCTCGCAAAGCT 44251

Qy        171 -PheAlaAspProProLysArgSerLeuThrArgSer----------------------- 182
Db      44250 GTTCGCAGACAA-CCTCGTCCACACCTCACCGCAAGTCATCGCCGATCATCCAGGGTGGC 44192

Qy        183 ------------------------AlaThrGlyProProIleHisGlyAlaAlaIlePr 194
Db      44191 TGTTCGCGCAGAGGTGTGAGGCGTGGTGGAAGGGCCA---------------------- 44155

Qy        194 oLeuLysPheProAspProValAsnThrGlySerLysArgProSerLeuGluSerGluAs 214
Db      44154 ------------------------ACTGGA---CGAAGACCCAGTCTCAGTTCTGCTCG 44123

Qy        214 nLeuAsnGlnCysThr---LysArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAl 233
Db      44122 TCTACCGCAGTCCTCGTGTAAAAGACACCTCACGACTTTATCG----------------- 44080

Qy        233 aAlaAlaProProValProIleAlaSerAlaLeuAspLysValProThrArgArgHisAl 253
Db      44079 -------GCCTGTGTCGAATTTGATTCTTTCCGCGAACAAAGTCGAGTACGTAGCACGGG 44027

Qy        253 aAsnThr------------ArgAspProThrAlaThrGlyHisArgArgAlaAspGlnVa 269
Db      44026 GAACTCGCTTTGATCCGACCATGATCCACTGCGACAAGGGCACCGA------------GT 43979

Qy        269 lAspSerPheAspThrSerGlnGlyThrSerTyrGlySerSerValPheSerAlaCysAr 289
Db      43978 GGATGGTGCTCAGGATGACTCAGGCTG-ACCAATTCCCCAAGGATCCGATTGACATGTCG 43920

Qy        289 gHisAsnGlnSerThrThrGlnSer-----------------------------SerPh 299
Db      43919 GAGGTCGCAAAACTGCTCAGTGTCGCTGTCGACGGGGCGCTGAACGAGACGATAGGCAAG 43860

Qy        299 eGluAlaProProSerGlnProArgGluLysArgProValAspAlaThrValPheGluAl 319
Db      43859 GGAAGGATGGCCGCTGCGTGACAGACACCAAGACATTGCTGCCGCTGCGAAGCGTCAGCC 43800

Qy        319 aGlyHisLeuIleGluSerProSerLysGlyArgThrThrLysSerHisIleAspAsn-- 338
Db      43799 AGTTGAACCAATACACACGTT-GCCCGCAGGCGTACAAGCTG-GCTCGCATCGACAAGGT 43742

Qy        339 -------GlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThrTyrTyrGl 356
Db      43741 ATGGGCTCGTCCTGCGGCGTGGTTGCCGCAGGGC---ACCGCGTTTCACACCGTCGCTGA 43685

Qy        356 uSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLe 376
```

```
Db  43684 GGTCTACGAGAAGGCTCTGGCCGAGGGC------------------CGCGAGATGAGCCT 43643

Qy    376 uAlaArgSerGluGluSerAlaArgSerGlnValGlnValHisAlaProValValAlaAl 396
Db  43642 CGAACGGGCACAAGAGATCTTCCGCGAGGAG---------TACGCCAAGGACATCGGAGC 43592

Qy    396 aArgLeuArgAsnIleTrpProLysPheProLysTrpLeuHisGluAlaProLeuAlaVa 416
Db  43591 A---CTGTGCGACGAAACCCCGAACTTCGAGTGGTGGTTTTGGTCCGGGCCGTACAACGG 43535

Qy    416 lAlaTrpGluValThrArgLeuPheMetHisCysLysValAspLeuGluAspGluSerLe 436
Db  43534 TGAGCGTGACATCGAGCGCCGGTTC---CACCTTGGACTGGAGCAGGTAGAGAAGTTCAT 43478

Qy    436 uGlyLeuLysTyrAspProSer---------TrpSerThr---AlaArgAspVal----- 450
Db  43477 CGCGTGGCGGAAGGACAAGGGCCAGCAGATCTGGACTACCCCTGGGCGAGGATCTGCTC 43418

Qy    451 -ThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArgGlyLysProPheProGl 470
Db  43417 TACGGAGGCGTGGAAAGACACCAACTGCAAGGAGTGTGAACAGCCCAAGCCGGCCATCGA 43358

Qy    470 uLysProProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSerLysGly-- 489
Db  43357 GCTTCCG-----------------------------TTCAACATCGAGCTGGACGGCAT 43328

Qy    490 ---------------------SerAlaValValLeuSerAla---------ValLeuAs 499
Db  43327 CCGCGTCCGAGGGTTCATCGACGCGGTCGTCGTGGTCAACGGTGAGCTTCGAGTCCGGGA 43268

Qy    499 pTyrAsnProAspAsnSerProThrAlaProLeuTyrLeuValLysLeuLysProLeuMe 519
Db  43267 CTACAAGACCGGCAACTCGCCGGGTGACGACTTCCAGCTCGGC--------------GT 43223

Qy    519 tPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIleLe 539
Db  43222 GTACGCGCTCGCGGTCGCGATGACC-----------------------TATGACGTGGA 43187

Qy    539 uIleProSerProThrSerThrSerProSerValProProValValSerLysGlnProGl 559
Db  43186 GGCTCCGAAGACGGGCGACTACTTCATGGCGGGGAAGAAGGGCATCAAGGCGAAGCCCAC 43127

Qy    559 yAlaValGluGluValIleGlnTrp------------LeuThrMetGlyGlnHisSerLe 575
Db  43126 TGCGCCCTACGACCTGACCGATTGGACGCGGGAGCGGATCACCGAGAGGTTCCATGAGGT 43067

Qy    575 uValGlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuAr 595
Db  43066 CGAGGCTCGG--------------------ATCCAGGCTGGTGACTTCGAGCCTCTG-- 43030

Qy    595 gGluPheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePh 615
Db  43029 ------------------------CCTGAACCT---------GACAAATGCGGCTTC-- 43006

Qy    615 eAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAl 635
Db  43005 -------TGTGATGTGAATTACAGCTGCCCTGTTTTTAAGTG-AGGTCA------CCTTG 42960

Qy    635 aGluGluProVal-----------------------------------------------Gl 640
Db  42959 ACAGCCACCGATACTCACCTTCCCTCTAAGGCAGGCACGAAGGAATACAAGAAAACGCTC 42900

Qy    640 uGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAspAsnAs 660
Db  42899 GAACTCGACCCCGACGCGGGTTACACGTACGTGGAAATGGGGCCGATTCCTGACAAGCCG 42840

Qy    660 nThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTy 680
Db  42839 AAG-TGG------CACCAGGAGCTGTACGCC------------------TCGCAGAAGTT 42805

Qy    680 rAlaIleMetThrLeuGlu---------------------ProHisGlnIleArgHisHi 693
Db  42804 TCCGTTCCCGACCCCAGAGGCTGCACGCAGGTTTGCCCAAGCGGAGGCCGTGCGTCATCC 42745

Qy    693 sLysThrAspLeuLeu-SerProSerGlyThrGlyGluValMetAsnAspGlyValGlyA 713
Db  42744 CGGACGAGAGATCGTGATCGAGTTCGAGGACG---------GTCGGAGATGGAACGGGCA 42694

Qy    713 rgMetSer---------ArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyA 730
Db  42693 GGAGTTCGTGACGCGAGAGGTATCGAGCCGAACATGAACGG------------------ 42653

Qy    730 spValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp- 749
Db  42652 --------CCGGCCACTTCATCCCGC----AGTACCAGGAAGAGTTCGGTCTGTGACC 42607

Qy    750 --ValAspAspThrGlyAspGluAsp---------TrpIleGluThrTyrProSerGlnA 766
Db  42606 AACTTGACAGCCACCGAGAGGAAGACATGACGAACTGGGATCCAAACCACCCGTCGCTGA 42547

Qy    766 rgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValA 786
Db  42546 GG------------TCACCTATAGCCCCGCACGAGACAGCCTGCGTCCTACGG------- 42506

Qy    786 laSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgA 806
Db  42505 --------ATGCACCGCGCTGGATAC-----------------------------AAGG 42484

Qy    806 laArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnA 826
Db  42483 GTCGGGCGACCATGAAGATCCTCAAGCTGCGCGGCACCCGGCTGATGAACCAGATGCAGC 42424

Qy    826 rgGlnPheSerGluGlnLysHisAla------LeuAsnArgProValGluPheArgGlnT 844
Db  42423 GAGCCTTGGACGCCGAGACGCACGCGGCTCGCGCCGGTCGGCCG---------------- 42380

Qy    844 rpValTyrGluSerTyr------------------------------------SerSerA 852
Db  42379 --ATTCACGACGCCTTGATCGATCCGAAGAAGGTGAAGTGAGGCCCAGCCTCTACGACCA 42322

Qy    852 rgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnG 872
Db  42321 GATCGTCGCAATCCTCGACCGGCACGATTCCTGCTGCGGATTCGAG-CGTGACGACGCGG 42263

Qy    872 luGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLys--------------- 886
Db  42262 CAGAGGAGCTGCACGCGCTGGTGATCGAGCTGTACGGCCCGCCGTTCTGAAAGGAAGCCA 42203

Qy    887 ----------------------------------LysGlnLysTyrLeuGlnAspIle- 894
Db  42202 TGAGCAACACCTGGTTCACGTCCGATCTACACATCGGACACAAGCGGCTGATGGAGATCC 42143

Qy    895 ---------------------------------------------AlaTrpAspLeuG 899
Db  42142 GCAACCTGGCGGATGATGTCGAAGAGCACGACGCCACCCTGGCGAAGGCGTGGGACTCGG 42083

Qy    899 lnLysArgLysCysAspThrLeu------------------------------LysS 908
Db  42082 TCGTCGGCAAGGACGACACGGTCTGGATCCTAGGTGACATCTCCTCTGGCTCCACCAAGG 42023

Qy    908 erLysLeuAsnIle-----------------ArgValGlyArgSerAlaTyrIleTyr- 921
Db  42022 GGCAGATACATGCCCTCGGGTGGATCTCGGATCGGCCAGGCCGCAAACGGCTGATCCTGG 41963

Qy    922 -----------------------------------------MetIleAlaAspPheT 927
Db  41962 GCAACCACGATGGGCCCCACCCGATGAACCGCGATGCTCACAAGCTGGTTGGTGCGTACT 41903

Qy    927 rpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArg------- 944
Db  41902 GGATGGTGTTCGAG------------CATGTGTCGACGGCAGCTCGTATCCGGGTGCCGC 41855

Qy    944 ----------------------------------------------------------- 944
Db  41854 TGTACGGGGACGCTGGCGGGCACACAGATGTTCTGCTGAGCCACTTCCCATACGTCGGGG 41795
```

```
Qy        945 --------------AspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuV 960
                            |||         :::  |||      |||      :::|||:
Db      41794 ACCACACCAGCGAGGACCGACACACCCAGTGGAGGCTCCGTGACGATGGCAAGATCCTGA 41735

Qy        960 alAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysP 980
              ::                |||       |||   :::         |||   :::   :::
Db      41734 TC---------CACGGGCACACGCATTCGCCGATGATCCTGTCCCGACACATCCACCGGA 41684

Qy        980 roGluLeuHis---------------------------------------SerLeuLysA 987
               ::::::|||                                        :::    |
Db      41683 GACAGATCCACGTGGGGATCGACGCCTGGGGCCGGCTGGTCTCTCGCGACGAGATCTATG 41624

Qy        987 spValIle-IlePheSerThrLysGlyAspVal---------------ProLeuAlaLys 1001
              ||:::::: |||   ||||||:::         |||                |||
Db      41623 ACCTGGTCAATCACATCCACGAGGAGGAAGGTGTACACACCTAGACAATCGCTCTACATC 41564

Qy       1002 LysLeuSerGlyGlyAsp------------------------------------------ 1007
              :::    |||   ||||||
Db      41563 CGAGGGTCGGCGGGGGATCCGCTACCCCGGTCTGGGCCGCGCTTGACCAGAAAGGTACA 41504

Qy       1008 ------TyrAspGlyAspMetAlaTrpValCysTrpAspProGluIle---ValAspGly 1024
                          |||   :::        ||||||        |||
Db      41503 CATCTCCGACGAGGCCAGCTAGTCCTGGTCTGCGCCGGCCCCGGCACCGGGAAGTCGGCG 41444

Qy       1025 PheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLys 1044
              ||||||    |||                           :::   |||        |||
Db      41443 TTCGTCCTCGCCTACGCGCTCAAGTCGAAGGTGCCGACGTTGTACTTCTCGGCCGACTCT 41384

Qy       1045 ThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySer------------------ 1058
                    |||   ||||||      |||         |||
Db      41383 GATGCGTTCACCCAGCTCTCCCGTTCGGTGTCCATCCTCAGCGGATGGTCGCTCGAACGA 41324

Qy       1059 ---------AlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSer--------- 1072
                          |||   :::||||||         |||   :::      |||
Db      41323 GCTACGCGGGCTGTTCGCGAGCAGAACATCGAGGACGCGGTAGCTGACTCACTAGACGAG 41264

Qy       1073 ------------PheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyr 1088
                          |||::::::         |||:::
Db      41263 ATCCCGATCCGGTTCAACTACAAGGCATCACCGTCGTTG------------GACGAGATC 41216

Qy       1089 LysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSer 1108
              :::       |||         :::           :::   ||||||:::|||:::
Db      41215 GAGAACGCTCTGGCCGCGTATGACGCGCTGTACGAGGACTTCCCAGCGTTGATCGTCGTG 41156

Qy       1109 SerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSer 1128
                    :::   |||:::         |||   :::|||
Db      41155 GACAACATCACCAACGTCCGCACCGATTCAAGCGAGGGA--------------------- 41117

Qy       1129 TrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyr 1148
                                 :::         |||      ||||||   |||   ::::::
Db      41116 ---------------GACGACCCGTTCTCGGGTCTGGAGTCGCTGATGGACTACCTACAC 41072

Qy       1149 LysSerAspSerTrpLeuGlyArg------------------------------------ 1156
              :::                :::||||||
Db      41071 GAG------------ATGGGCCGTGAGACAGGCTCATGCGTCGTCGGGCTCCACCACGTC 41024

Qy       1157 ------------------GlyGluProThrHisIleIleAspTyrLeuLysPheSerIle 1170
                                |||:::            :::      :::|||      |||
Db      41023 ACCGGCCCGCACAACGACGGCGACAAGCCGATCCCGTTGTCGGGCATCAAAGGTCAGATC 40964

Qy       1171 AlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLys 1190
                 |||      :::   :::   :::            |||
Db      40963 GGGCGT------GTTCCTGAGATGATCCTCACTCTCCACCGA------------------ 40928

Qy       1191 AspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePhe 1210
                          ||||||            :::   ||||||:::   :::         |||
Db      40927 ---GTATCGGACGGG---------TTCGGCCCGGACATGCTCAACGTCTCCACG---GTC 40883

Qy       1211 LysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLys------ 1228
              |||               ||||||      |||         |||   :::|||:::
Db      40882 AAGAACCGAGGGGGGAAGTCGGACCCGTCGGGACAAGACTTCGCTTCCTTGGAGTTCGTC 40823

Qy       1229 ------------------AsnArgIleGly------------------------------ 1232
                                |||   |||
Db      40822 GGAGACACTATGCAGATCAACGACTTCGGTCACTAACTTGACATCCACCAAACAGAAAGT 40763

Qy       1233 ---GluValGluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSerLys 1251
                 |||   |||:::::::::                  :::::::::   |||
Db      40762 AGCGAACATGAAGAAGATCAT------------CGCCACCGCACTCATCGCCGGATCGGC 40715

Qy       1252 AspProTyrProValArgVal 1258
              ||||||   |||:::||||||
Db      40714 GATCCTGGGCCTCTCCGCGTG 40694
```

```
RESULT 15
US-08-705-557-1/c
; Sequence 1, Application US/08705557
; Patent No. 6300061
;  GENERAL INFORMATION:
;    APPLICANT:  William R. Jacobs, Jr.
;    APPLICANT:  Barry R. Bloom
;    APPLICANT:  Graham F. Hatfull
;    TITLE OF INVENTION:  MYCOBACTERIAL SPECIES-SPECIFIC
;    TITLE OF INVENTION:  REPORTER MYCOBACTERIOPHAGES
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amster, Rothstein & Ebenstein
;      STREET:  90 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch 1.44 Mb storage diskette
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Processor (ASCII)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/705,557
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/057,531
;      FILING DATE:
;      APPLICATION NUMBER:  07/833,431
;      FILING DATE:  February 7, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pasqualini, Patricia A.
;      REGISTRATION NUMBER:  34,894
;      REFERENCE/DOCKET NUMBER:  96700/238
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  52297
;      TYPE:  nucleotide
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  phage genome sequence
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    FRAGMENT TYPE:  not applicable.
;    ORIGINAL SOURCE:
;      ORGANISM:  mycobacteriophage L5
;      STRAIN:  not applicable
;      INDIVIDUAL ISOLATE:  L5
;      DEVELOPMENTAL STAGE:  not applicable
;      HAPLOTYPE:  not applicable
;      TISSUE TYPE:  not applicable
;      CELL TYPE:  not applicable
;      CELL LINE:  not applicable
;      ORGANELLE:  not applicable
```

```
;     IMMEDIATE SOURCE:  mycobacteriophage L5 particles
;     POSITION IN GENOME:  entire genome
;     FEATURE:
;       NAME/KEY:
;       LOCATION:
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
;     PUBLICATION INFORMATION:
;       AUTHORS:  Hatfull and Sarkis
;       TITLE:  DNA Sequence, Structure and Gene
;       TITLE:  Expression of Mycobacteriophage L5:
;       TITLE:  A Phage System for Mycobacterial
;       TITLE:  Genetics
;       JOURNAL:  Molecular Microbiology
;       VOLUME:  7
;       PAGES:  395-405
;       DATE:  1993
US-08-705-557-1

Alignment Scores:
Pred. No.:              0.00719        Length:         52297
Score:                  171.00         Matches:        308
Percent Similarity:     32.70%         Conservative:   192
Best Local Similarity:  20.14%         Mismatches:     524
Query Match:            2.31%          Indels:         513
DB:                     4              Gaps:           76

US-09-913-878A-2 (1-1402) x US-08-705-557-1 (1-52297)

Qy         35 ThrLeuThrProHisArgArgLysGluLeuAlaGluSerAspGluAspPheGlyArgHis 54
                 |||     ||||||    :::    :::    :::         |||   |||   :::
Db      44666 ACGCGGAAACCTCACGATAAGTCGAAGGCTTTTCAGAGGCTCCGGGATCACGGCGAGTAC 44607

Qy         55 --------------------------------AspLysIleTyrArgAlaLeuAsnPhe 63
                                              ||||||    |||.   |||:::::::
Db      44606 ATGCGAGTGGGTAGCTTCCGCACGATGTCCGAAGACAAGAAGAAGCGCCTATTGTCGTGG 44547

Qy         64 LeuTyrTrpArgLysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGlu--- 82
                          :::||||||   |||        |||   :::    ||||||
Db      44546 TGGAAGATGCTCCGCGACGACGACTTA---GTGCTCGAGTTCGACCCATCCATCGAGCCC 44490

Qy         83 ---------AlaLysAlaAlaSerSerAsnTrpValProLysAlaHisAlaAspProAsp 99
                                       :::||||||:::       |||       |||
Db      44489 TACGAAGGTATGGCCGGCGGCGGGTTCAGGTACGTCCCCCGAGGCATCGAGGATGACGAT 44430

Qy        100 ThrLeu-----------ProTrpSerLys-GluProProArgAlaAlaThrAlaGlyGl 115
                 |||                          ||||||||       ::::::|||::
Db      44429 CTCCTGATCCGGGTGAACGAGCACACCAACCTCACCGCCGAGGGTGAACTCCTCTGGTCG 44370

Qy        115 nGlnTrpAlaLeuGlnThrValLeuLeuGluValLeuAsnArgPheMetProProProAs 135
              :
Db      44369 T---------------------------------------------------------GG 44367

Qy        135 nAsnThrProGlyArgThrPheGlyArgThrLeuSerGlyProSerGlyLeuSerArgPr 155
                    |||      |||:::|||   |||:::                |||      ||
Db      44366 CCAGACGACATCGAGGAGCTTCTCTCGGAGCCCTAAAAGTCACGACCGGTTGTGTGAGCC 44307

Qy        155 oThrSerThrAsnThrLysArgLysAspGluProAlaAsnValThr-------------- 170
              ||||                 :::||||||:::   |||       :::|||
Db      44306 AACCCAGGCC---GCGCAACGAAAGGAAACCCCGTGCCC-CTCACTCGGCTCGCAAAGCT 44251

Qy        171 -PheAlaAspProProLysArgSerLeuThrArgSer----------------------- 182
               ||||||||    |||:::      ||||||   |||
Db      44250 GTTCGCAGACAA-CCTCGTCCACACCTCACCGCAAGTCATCGCCGATCATCCAGGGTGGC 44192

Qy        183 ------------------------AlaThrGlyProProIleHisGlyAlaAlaIlePr 194
                                              ||||||
Db      44191 TGTTCGCGCAGAGGTGTGAGGCGTGGTGGAAGGGCCA----------------------- 44155

Qy        194 oLeuLysPheProAspProValAsnThrGlySerLysArgProSerLeuGluSerGluAs 214
                                       ||||||   :::|||||||||||||   |||
Db      44154 -------------------------ACTGGA---CGAAGACCCAGTCTCAGTTCTGCTCG 44123

Qy        214 nLeuAsnGlnCysThr---LysArgAlaLysGlyLysLeuSerAspAsnValAlaAlaAl 233
               |||   |||    :::    ||||||                  ||||||
Db      44122 TCTACCGCAGTCCTCGTGTAAAAGACACCTCACGACTTTATCG----------------- 44080

Qy        233 aAlaAlaProProValProIleAlaSerAlaLeuAspLysValProThrArgArgHisAl 253
                                  |||        |||      ::::::         |||
Db      44079 -------GCCTGTGTCGAATTTGATTCTTTCCGCGAACAAAGTCGAGTACGTAGCACGGG 44027

Qy        253 aAsnThr-----------ArgAspProThrAlaThrGlyHisArgArgAlaAspGlnVa 269
               |||:::                    |||||         |||||||||        ||
Db      44026 GAACTCGCTTTGATCCGACCATGATCCACTGCGACAAGGGCACCGA-----------GT 43979

Qy        269 lAspSerPheAspThrSerGlnGlyThrSerTyrGlySerSerValPheSerAlaCysAr 289
               ||||                     |||    :::              :::      |||||
Db      43978 GGATGGTGCTCAGGATGACTCAGGCTG-ACCAATTCCCCAAGGATCCGATTGACATGTCG 43920

Qy        289 gHisAsnGlnSerThrThrGlnSer----------------------------SerPh 299
              |   :::|||:::   :::   |||                                   :::
Db      43919 GAGGTCGCAAAACTGCTCAGTGTCGCTGTCGACGGGGCGCTGAACGAGACGATAGGCAAG 43860

Qy        299 eGluAlaProProSerGlnProArgGluLysArgProValAspAlaThrValPheGluAl 319
                |||      |||      :::    |||      :::         |||       :::
Db      43859 GGAAGGATGGCCGCTGCGTGACAGACACCAAGACATTGCTGCCGCTGCGAAGCGTCAGCC 43800

Qy        319 aGlyHisLeuIleGluSerProSerLysGlyArgThrThrLysSerHisIleAspAsn-- 338
                       |||   :::   ::::::   ||||||:::   :::   ||||||
Db      43799 AGTTGAACCAATACACACGTT-GCCCGCAGGCGTACAAGCTG-GCTCGCATCGACAAGGT 43742

Qy        339 -------GlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThrTyrTyrGl 356
                     :::|||   :::                ||||||   |||:::|||   |||      ||
Db      43741 ATGGGCTCGTCCTGCGGCGTGGTTGCCGCAGGGC---ACCGCGTTTCACACCGTCGCTGA 43685

Qy        356 uSerPheProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLe 376
              |   :::      :::      ||||||                  |||       |||
Db      43684 GGTCTACGAGAAGGCTCTGGCCGAGGGC------------------CGCGAGATGAGCCT 43643

Qy        376 uAlaArgSerGluGluSerAlaArgSerGlnValGlnValHisAlaProValValAlaAl 396
              |   |||::::::|||      |||   :::            :::|||      :::   ||
Db      43642 CGAACGGGCACAAGAGATCTTCCGCGAGGAG---------TACGCCAAGGACATCGGAGC 43592

Qy        396 aArgLeuArgAsnIleTrpProLysPheProLysTrpLeuHisGluAlaProLeuAlaVa 416
              |   |||   :::      |||   |||         |||            |||
Db      43591 A---CTGTGCGACGAAACCCCGAACTTCGAGTGGTGGTTTTGGTCCGGGCCGTACAACGG 43535

Qy        416 lAlaTrpGluValThrArgLeuPheMetHisCysLysValAspLeuGluAspGluSerLe 436
                       ::::::   |||   |||   |||         ::::::         ::::::   ::
Db      43534 TGAGCGTGACATCGAGCGCCGGTTC---CACCTTGGACTGGAGCAGGTAGAGAAGTTCAT 43478

Qy        436 uGlyLeuLysTyrAspProSer---------TrpSerThr---AlaArgAspVal----- 450
              :     :::   |||                   |||:::|||      |||   :::
Db      43477 CGCGTGGCGGAAGGACAAGGGCCAGCAGATCTGGACTACCCCTGGGCGAGGGATCTGCTC 43418

Qy        451 -ThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArgGlyLysProPheProGl 470
               |||:::   ||||||               :::      :::   ||||||      ||
Db      43417 TACGGAGGCGTGGAAAGACACCAACTGCAAGGAGTGTGAACAGCCCAAGCCGGCCATCGA 43358

Qy        470 uLysProProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSerLysGly-- 489
                 |||                                |||   |||   |||
Db      43357 GCTTCCG-----------------------------TTCAACATCGAGCTGGACGGCAT 43328

Qy        490 ---------------------SerAlaValValLeuSerAla---------ValLeuAs 499
                                   :::   ||||||::::::         |||   ||
Db      43327 CCGCGTCCGAGGGTTCATCGACGCGGTCGTCGTGGTCAACGGTGAGCTTCGAGTCCGGGA 43268

Qy        499 pTyrAsnProAspAsnSerProThrAlaProLeuTyrLeuValLysLeuLysProLeuMe 519
              ||||         |||||||||               |||                  ::
Db      43267 CTACAAGACCGGCAACTCGCCGGGTGACGACTTCCAGCTCGGC--------------GT 43223

Qy        519 tPheGluGlnGlyCysArgLeuThrArgArgPheGlyProAspArgPhePheGluIleLe 539
              ::::                   :::|||                        :::::::::
Db      43222 GTACGCGCTCGCGGTCGCGATGACC-----------------------TATGACGTGGA 43187

Qy        539 uIleProSerProThrSerThrSerProSerValProProValValSerLysGlnProGl 559
```

```
Db     43186 GGCTCCGAAGACGGGCGACTACTTCATGGCGGGGAAGAAGGGCATCAAGGCGAAGCCCAC 43127
Qy       559 yAlaValGluGluValIleGlnTrp------------LeuThrMetGlyGlnHisSerLe 575
Db     43126 TGCGCCCTACGACCTGACCGATTGGACGCGGGAGCGGATCACCGAGAGGTTCCATGAGGT 43067
Qy       575 uValGlyArgGlnTrpArgAlaPhePheAlaLysAspAlaGlyTyrArgLysProLeuAr 595
Db     43066 CGAGGCTCGG---------------------ATCCAGGCTGGTGACTTCGAGCCTCTG-- 43030
Qy       595 gGluPheGlnLeuArgAlaGluAspProLysProIleIleLysGluArgValHisPhePh 615
Db     43029 --------------------------CCTGAACCT----------GACAAATGCGGCTTC-- 43006
Qy       615 eAlaGluThrGlyIleThrPheArgProAspValPheLysThrArgSerValValProAl 635
Db     43005 -------TGTGATGTGAATTACAGCTGCCCTGTTTTTAAGTG-AGGTCA------CCTTG 42960
Qy       635 aGluGluProVal---------------------------------------------Gl 640
Db     42959 ACAGCCACCGATACTCACCTTCCCTCTAAGGCAGGCACGAAGGAATACAAGAAAACGCTC 42900
Qy       640 uGlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeuAspAsnAs 660
Db     42899 GAACTCGACCCCGACGCGGGTTACACGTACGTGGAAATGGGGCCGATTCCTGACAAGCCG 42840
Qy       660 nThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGlyLeuSerLysThrTy 680
Db     42839 AAG-TGG------CACCAGGAGCTGTACGCC-----------------TCGCAGAAGTT 42805
Qy       680 rAlaIleMetThrLeuGlu--------------------ProHisGlnIleArgHisHi 693
Db     42804 TCCGTTCCCGACCCCAGAGGCTGCACGCAGGTTTGCCCAAGCGGAGGCCGTGCGTCATCC 42745
Qy       693 sLysThrAspLeuLeu-SerProSerGlyThrGlyGluValMetAsnAspGlyValGlyA 713
Db     42744 CGGACGAGAGATCGTGATCGAGTTCGAGGACG---------GTCGGAGATGGAACGGGCA 42694
Qy       713 rgMetSer---------ArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyA 730
Db     42693 GGAGTTCGTGACGCGAGAGGTATCGAGCCGAACATGAACGG------------------- 42653
Qy       730 spValProSerAlaValGlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAsp- 749
Db     42652 --------CCGGCCACTTCATCCCGC------AGTACCAGGAAGAGTTCGGTCTGTGACC 42607
Qy       750 --ValAspAspThrGlyAspGluAsp---------TrpIleGluThrTyrProSerGlnA 766
Db     42606 AACTTGACAGCCACCGAGAGGAAGACATGACGAACTGGGATCCAAACCACCCGTCGCTGA 42547
Qy       766 rgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSerValA 786
Db     42546 GG------------TCACCTATAGCCCCGCACGAGACAGCCTGCGTCCTACGG------- 42506
Qy       786 laSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgA 806
Db     42505 --------ATGCACCGCGCTGGATAC-----------------------------AAGG 42484
Qy       806 laArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeuGlnA 826
Db     42483 GTCGGGCGACCATGAAGATCCTCAAGCTGCGCGGCACCCGGCTGATGAACCAGATGCAGC 42424
Qy       826 rgGlnPheSerGluGlnLysHisAla------LeuAsnArgProValGluPheArgGlnT 844
Db     42423 GAGCCTTGGACGCCGAGACGCACGCGGCTCGCGCCGGTCGGCCG---------------- 42380
Qy       844 rpValTyrGluSerTyr------------------------------------SerSerA 852
Db     42379 --ATTCACGACGCCTTGATCGATCCGAAGAAGGTGAAGTGAGGCCCAGCCTCTACGACCA 42322
Qy       852 rgAlaThrArgValSerHisGlyArgValProPheLeuAlaGlyLeuProAspSerGlnG 872
Db     42321 GATCGTCGCAATCCTCGACCGGCACGATTCCTGCTGCGGATTCGAG-CGTGACGACGCGG 42263
Qy       872 luGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLys--------------- 886
Db     42262 CAGAGGAGCTGCACGCGCTGGTGATCGAGCTGTACGGCCCGCCGTTCTGAAAGGAAGCCA 42203
Qy       887 -----------------------------------LysGlnLysTyrLeuGlnAspIle- 894
Db     42202 TGAGCAACACCTGGTTCACGTCCGATCTACACATCGGACACAAGCGGCTGATGGAGATCC 42143
Qy       895 ---------------------------------------------AlaTrpAspLeuG 899
Db     42142 GCAACCTGGCGGATGATGTCGAAGAGCACGACGCCACCCTGGCGAAGGCGTGGGACTCGG 42083
Qy       899 lnLysArgLysCysAspThrLeu--------------------------------LysS 908
Db     42082 TCGTCGGCAAGGACGACACGGTCTGGATCCTAGGTGACATCTCCTCTGGCTCCACCAAGG 42023
Qy       908 erLysLeuAsnIle------------------ArgValGlyArgSerAlaTyrIleTyr- 921
Db     42022 GGCAGATACATGCCCTCGGGTGGATCTCGGATCGGCCAGGCCGCAAACGGCTGATCCTGG 41963
Qy       922 -----------------------------------------MetIleAlaAspPheT 927
Db     41962 GCAACCACGATGGGCCCCACCCGATGAACCGCGATGCTCACAAGCTGGTTGGTGCGTACT 41903
Qy       927 rpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArg------- 944
Db     41902 GGATGGTGTTCGAG------------CATGTGTCGACGGCAGCTCGTATCCGGGTGCCGC 41855
Qy       944 ------------------------------------------------------------ 944
Db     41854 TGTACGGGGACGCTGGCGGGCACACAGATGTTCTGCTGAGCCACTTCCCATACGTCGGGG 41795
Qy       945 --------------AspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuV 960
Db     41794 ACCACACCAGCGAGGACCGACACACCCAGTGGAGGCTCCGTGACGATGGCAAGATCCTGA 41735
Qy       960 alAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysP 980
Db     41734 TC---------CACGGGCACACGCATTCGCCGATGATCCTGTCCCGACACATCCACCGGA 41684
Qy       980 roGluLeuHis-----------------------------------------SerLeuLysA 987
Db     41683 GACAGATCCACGTGGGGATCGACGCCTGGGGCCGGCTGGTCTCTCGCGACGAGATCTATG 41624
Qy       987 spValIle-IlePheSerThrLysGlyAspVal---------------ProLeuAlaLys 1001
Db     41623 ACCTGGTCAATCACATCCACGAGGAGGAAGGTGTACACACCTAGACAATCGCTCTACATC 41564
Qy      1002 LysLeuSerGlyGlyAsp------------------------------------------ 1007
Db     41563 CGAGGGTCGGCGGGGGATCCGCTACCCCCGGTCTGGGCCGCGCTTGACCAGAAAGGTACA 41504
Qy      1008 ------TyrAspGlyAspMetAlaTrpValCysTrpAspProGluIle---ValAspGly 1024
Db     41503 CATCTCCGACGAGGCCAGCTAGTCCTGGTCTGCGCCGGCCCCGGCACCGGGAAGTCGGCG 41444
Qy      1025 PheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLys 1044
Db     41443 TTCGTCCTCGCCTACGCGCTCAAGTCGAAGGTGCCGACGTTGTACTTCTCGGCCGACTCT 41384
Qy      1045 ThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySer------------------ 1058
Db     41383 GATGCGTTCACCCAGCTCTCCCGTTCGGTGTCCATCCTCAGCGGATGGTCGCTCGAACGA 41324
Qy      1059 ---------AlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSer--------- 1072
Db     41323 GCTACGCGGGCTGTTCGCGAGCAGAACATCGAGGACGCGGTAGCTGACTCACTAGACGAG 41264
Qy      1073 ------------PheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyr 1088
Db     41263 ATCCCGATCCGGTTCAACTACAAGGCATCACCGTCGTTG------------GACGAGATC 41216
```

```
Qy     1089 LysGluArgLeuCysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSer 1108
            :::      |||         :::            :::   ||||||:::|||:::
Db    41215 GAGAACGCTCTGGCCGCGTATGACGCGCTGTACGAGGACTTCCCAGCGTTGATCGTCGTG 41156

Qy     1109 SerLeuValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSer 1128
                  :::   |||:::         |||   :::|||
Db    41155 GACAACATCACCAACGTCCGCACCGATTCAAGCGAGGGA--------------------- 41117

Qy     1129 TrpAlaGlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMetTyr 1148
                              :::         |||      ||||||   |||   ::::::
Db    41116 ---------------GACGACCCGTTCTCGGGTCTGGAGTCGCTGATGGACTACCTACAC 41072

Qy     1149 LysSerAspSerTrpLeuGlyArg------------------------------------ 1156
            :::            :::||||||
Db    41071 GAG------------ATGGGCCGTGAGACAGGCTCATGCGTCGTCGGGCTCCACCACGTC 41024

Qy     1157 ------------------GlyGluProThrHisIleIleAspTyrLeuLysPheSerIle 1170
                              |||:::            :::      :::|||      |||
Db    41023 ACCGGCCCGCACAACGACGGCGACAAGCCGATCCCGTTGTCGGGCATCAAAGGTCAGATC 40964

Qy     1171 AlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAlaMetLysAlaAlaLys 1190
               |||      :::   :::   :::         |||
Db    40963 GGGCGT------GTTCCTGAGATGATCCTCACTCTCCACCGA------------------ 40928

Qy     1191 AspThrGluAspGlyAlaHisPheTrpAspProAspLeuAlaSerTyrTyrThrPhePhe 1210
                     ||||||         :::   ||||||:::   :::      |||
Db    40927 ---GTATCGGACGGG---------TTCGGCCCGGACATGCTCAACGTCTCCACG---GTC 40883

Qy     1211 LysGluIleSerAspLysSerArgSerSerAlaLeuLeuPheThrThrLeuLys------ 1228
            |||            ||||||      |||         |||   :::|||:::
Db    40882 AAGAACCGAGGGGGGAAGTCGGACCCGTCGGGACAAGACTTCGCTTCCTTGGAGTTCGTC 40823

Qy     1229 ------------------AsnArgIleGly------------------------------ 1232
                              |||      |||
Db    40822 GGAGACACTATGCAGATCAACGACTTCGGTCACTAACTTGACATCCACCAAACAGAAAGT 40763

Qy     1233 ---GluValGluLysGluTyrGlyArgLeuValLysAsnLysGluMetArgAspSerLys 1251
               |||   |||:::::::::            :::::::::      |||
Db    40762 AGCGAACATGAAGAAGATCAT------------CGCCACCGCACTCATCGCCGGATCGGC 40715

Qy     1252 AspProTyrProValArgVal 1258
            ||||||   |||:::||||||
Db    40714 GATCCTGGGCCTCTCCGCGTG 40694
```

Search completed: March 31, 2004, 13:11:52
Job time : 675.141 secs

OM protein - protein search, using sw model

Run on: March 29, 2004, 08:42:26 ; Search time 28 Seconds
(without alignments)
4816.447 Million cell updates/sec

Title: US-09-913-878A-2
Perfect score: 7397
Sequence: 1 MNPITPRKRNSPVEEIINRL..........YEVLGDDDFDGIGFTGNGDY 1402

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	328	4.4	1116	2	T30828	RNA-directed RNA p
2	297	4.0	1114	2	T30819	RNA-directed RNA p
3	296.5	4.0	1780	2	T20695	hypothetical prote
4	279.5	3.8	1133	2	T01920	probable RNA-direc
5	265.5	3.6	1215	2	T11660	probable RNA-direc
6	232.5	3.1	3228	2	T21381	hypothetical prote
7	191	2.6	5105	2	T32650	hypothetical prote
8	182	2.5	1075	2	T38253	hypothetical prote
9	177.5	2.4	966	2	F84582	hypothetical prote
10	174	2.4	1226	2	S15053	hypothetical prote
11	169.5	2.3	905	2	G84582	hypothetical prote
12	163.5	2.2	4377	2	A55575	ankyrin 3, long sp
13	160.5	2.2	1024	2	T30868	RhoA-binding prote
14	160	2.2	1176	2	JN0583	myosin-light-chain
15	159	2.1	929	2	H84582	hypothetical prote
16	158.5	2.1	1736	2	A47747	tight junction pro
17	153	2.1	1388	2	T00063	hypothetical prote
18	153	2.1	1697	2	T00079	hypothetical prote
19	148.5	2.0	1418	2	S64918	hypothetical prote
20	148.5	2.0	1743	2	T15893	hypothetical prote
21	144	1.9	2224	1	KFHU5	coagulation factor
22	143.5	1.9	848	2	S48273	probable transcrip
23	143.5	1.9	1316	2	T00381	KIAA0633 protein -
24	143	1.9	967	2	T20336	hypothetical prote
25	143	1.9	1285	2	T14171	ataxin-2 - mouse
26	142.5	1.9	860	2	JC4566	chitinase (EC 3.2.
27	142.5	1.9	3942	2	T42730	Bassoon protein -
28	142	1.9	2409	1	A60979	versican precursor
29	141.5	1.9	1211	2	T42230	AF4 protein - mous
30	141.5	1.9	3660	1	S02041	dystrophin, muscle
31	141	1.9	1576	2	S65774	homeotic protein H
32	140.5	1.9	1228	2	T18897	hypothetical prote
33	140.5	1.9	1516	2	F83085	conserved hypothet
34	140.5	1.9	3587	2	I40486	surfactin syntheta
35	140.5	1.9	3968	2	A44265	trithorax homolog
36	139.5	1.9	1207	2	T00378	KIAA0641 protein -
37	139.5	1.9	1367	1	S48478	glucan 1,4-alpha-g
38	139	1.9	2437	2	S53611	MIBP1 protein - ra
39	139	1.9	3685	1	A27605	dystrophin, muscle
40	138.5	1.9	990	2	I51618	nucleolar phosphop
41	138	1.9	2910	2	T42214	otogelin - mouse
42	137	1.9	653	2	E84682	hypothetical prote
43	137	1.9	2700	2	D88450	protein F21H11.2 [
44	136.5	1.8	1403	2	T11583	probable translati
45	136.5	1.8	1469	2	T19168	hypothetical prote

ALIGNMENTS

RESULT 1
T30828
RNA-directed RNA polymerase - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 20-Jun-2000
C;Accession: T30828
R;Schiebel, W.; Pelissier, T.; Riedel, L.; Thalmeir, S.; Schiebel, R.; Kempe, D.; Lotts
submitted to the EMBL Data Library, October 1998
A;Description: Isolation of an RNA-directed RNA polymerase-specific cDNA clone from Ton
A;Reference number: Z20890
A;Accession: T30828
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1116 <SCH>
A;Cross-references: EMBL:AJ011576; PIDN:CAA09697.1
C;Genetics:
A;Note: RdRP
C;Superfamily: Arabidopsis probable RNA-directed RNA polymerase

```
  Query Match             4.4%;  Score 328;  DB 2;  Length 1116;
  Best Local Similarity   19.9%;  Pred. No. 1e-11;
  Matches  211;  Conservative  149;  Mismatches  386;  Indels  314;  Gaps   42;

Qy        441 DPSWSTARDVTDIWKTL-----------YRLDAFRGKPFPEKPPNDVFV--TAMTGNFES 487
              |   |     |  |  |            :|    |    |   |  :  |:   |  |:|
Db        193 DDQWVRTTDFTPSWIGLSSSLCLEFRNGVQLPNFSESFFYYKESMNQFILQTGFTFSFSQ 252

Qy        488 KGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGC--RLTRRFGPDRFFEILIPSPTS 545
              | : |  :        : |:          | |:  |: : ||    |     ||:::  |   :
Db        253 KLALVPIV----HLPEGIELPYKILFKISSLI-QHGCFPGLALNF---NFFQLVDPRRRN 304

Qy        546 TSPSVPPVVSKQPGAVEEV----------IQWLT-MGQHSLVGRQ--------------- 579
                  ::     |:|::              ::|||     |  |||
Db        305 --------IACIEHALEKLYYLKECCYDPVRWLTEQYDEYLKGRQLPKSPPITLDDGLVY 356

Qy        580 --------WRAFFAKDAGYRKPLREFQLR--AEDPKPIIK--------ERVHFF-----A 616
                      : :|       |      :     ||   :||      ::           |::|      |
Db        357 VRRVVVTPCKVYF---CGPEVNVSNRVLRNYSEDINNFLRVSFVDEEWEKIHSTDLLPRA 413

Qy        617 ETGITFRPDVFK-----TRSVVPAEEPVEQRTEFKVSQMLD---W-------LLQLDNNT 661
              ||   |  |:::            |:        :    ||:  |      |      |   |   |
Db        414 STGNGTRTDIYERILSTLRNGFIIGDKRFEFLAFSSSQLRDNSVWMFASRPGLTANDIRT 473

Qy        662 WQPHLKLFSRI-----QLGLSKTYAIMTLEPHQIRHHKTDLL-----SPSGTGEVMNDGV 711
              |      :     :       :||  |     :   ||     :    |: :::        |   ||    : :||:
Db        474 WMGDFRQIRNVAKYAARLGQSFGSSRETL---SVGRHEVEVIPDVACSLHGTNYIFSDGI 530

Qy        712 GRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECD 771
              |::|   |  |:      ||     ||:  |  |:|   ||:  :|               |  |  :
Db        531 GKISADFAHRVTIKCGLQYTPSSFQIRYGGYKGVVAVD-----------PYSSMK----- 574

Qy        772 FVDKHQRTLEVRSVASELKSAGLNLQLL------PVLEDRARDKVKMRQAIGDRLINDLQ 825
```

```
Db        575 --------LSLRKSMLKYESNNIKLDVLGWSKYQPCYLNRQLVTLLSTLGVIDDVLEQKQ 626

Qy        826 RQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDP 885

Db        627 NEAVDQLDAILHDSLKAQEALELMS----------------PGENTNILKEMLNCGYMP 669

Qy        886 KKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSS---- 941

Db        670 DAEPFLSMMLQTFRASKLLDLRTRTRIFIPNGRTMMGCLDESRTLEYGQVFVQFSGAGRR 729

Qy        942 KFRDEEESF--TLLSDCD------VLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFST 993

Db        730 QFFEESHPFNDSGSANCDFILKGNVVVAKNPCLHPGDIRVLRAVDVPALHHMVDCVVFPQ 789

Qy        994 KGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMAS 1053

Db        790 KGKRPHPNECSGSDLDGDIYFVCWDPDLIPPRQVQSMDYTPAPTTQLDHD---------- 839

Qy       1054 HGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNSVS------------- 1100

Db        840 ----------VTIEEVEEYF-----------TNY-----IINDSLGIIANAHVVFADRE 872

Qy       1101 -----NKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPD-PMYKSDSWL 1154

Db        873 PDMAMSDPCKQLAQLFSIAVDFPKTGVPAEIPS--QLRPKEYPDFMEKPDKPTYPSEKVI 930

Qy       1155 GRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEIS 1214

Db        931 GK--------------------------LFQKVKNK-------TPQASSIATFTRDVA 955

Qy       1215 DKSRSSALLFTTLKNRIGEV---EKEYGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPE 1271

Db        956 RKSYDSDMIVDGFEDYIDEAFYYKSEY-----DNKLGNLMDYYGIKTE------AEILSG 1004

Qy       1272 AMDKSGANYDSK--------VIRLLELSFLADREMNTW---------ALLRASTAFKLYY 1314

Db       1005 GIMKASKTFDRRKDAEAIGVAVRCLR------KEARAWFKRRSDIDDMLAKASAWYHVTY 1058

Qy       1315 HKS-------------PKFVWQMAGRQLAYIKAQMTSRP 1340

Db       1059 HHTYWGLYNEGLKRDHFISFPWCVYDQLIQIKKAKARKRP 1098
```

RESULT 2
T30819
RNA-directed RNA polymerase (EC 2.7.7.48) - tomato
C;Species: Lycopersicon esculentum (tomato)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T30819
R;Schiebel, W.; Pelissier, T.; Riedel, L.; Thalmeir, S.; Schiebel, R.; Kempe, D.; Lottsp
Plant Cell 10, 2087-2102, 1998
A;Title: Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato.
A;Reference number: Z20885; MUID:99055198; PMID:9836747
A;Accession: T30819
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1114 <SCH>
A;Cross-references: EMBL:Y10403; NID:g4038591; PIDN:CAA71421.1; PID:g4038592
C;Genetics:
A;Note: RdRP
C;Superfamily: Arabidopsis probable RNA-directed RNA polymerase
C;Keywords: nucleotidyltransferase

```
  Query Match             4.0%;  Score 297;  DB 2;  Length 1114;
  Best Local Similarity   20.4%;  Pred. No. 8.1e-10;
  Matches  175;  Conservative  133;  Mismatches  323;  Indels  228;  Gaps   32;

Qy        562 EEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGIT 621

Db        386 EDIDNFLRV---SFVDEEWEKLYSTDL-------------LPK----------ASTGSG 418

Qy        622 FRPDVFK------TRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTW----QPHLKL--- 668

Db        419 VRTNIYERILSTLRKGFVIGDKKFE-FLAFSSSQL------RDNSVWMFASRPGLTANDI 471

Qy        669 ------FSRI--------QLGLSKTYAIMTLEPHQIRHHKTDLL---SPSGTGEVMNDGV 711

Db        472 RAWMGDFSQIKNVAKYAARLGQSFGSSRETL---SVLRHEIEVIPDVKVHGTSYVFSDGI 528

Qy        712 GRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECD 771

Db        529 GKISGDFAHRVASKCGLQYTPSAFQIRYGGYKGVVGVDPDSSMK---LSLRKSMSKYESD 585

Qy        772 FVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQ 831

Db        586 NI-----KLDVLG-WSKYQPCYLNRQLITLLS---------TLGVKDEVLEQKQKEAVDQ 630

Qy        832 KHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYL 891

Db        631 LDAILHDSLKAQEALELMS----------------PGENTNILKAMLNCGYKPDAEPFL 673

Qy        892 QDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFS----SKFRDEE 947

Db        674 SMMLQTFRASKLLDLRTRSRIFIPNGRTMMGCLDESRTLEYGQVFVQFTGAGHGEFSDDL 733

Qy        948 ESF--------TLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPL 999

Db        734 HPFNNSRSTNSNFILKGNVVVAKNPCLHPGDIRVLKAVNVRALHHMVDCVVFPQKGKRPH 793

Qy       1000 AKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSA 1059

Db        794 PNECSGSDLDGDIYFVCWDQDMIPPRQVQPMEYPPAPSIQLDHD--------------- 837

Qy       1060 AKEQTTYDMIQKSFHFALQPNFLGMCTN----YKERLCYINNSVSNKPAIILSSLVGNLV 1115

Db        838 ----VTIEEVEEYFTNYIVNDSLGIIANAHVVFADR---EPDMAMSDPCKKLAELFSIAV 890

Qy       1116 DQSKQGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAI 1175

Db        891 DFPKTGVPAEIPS--QLRPK------EYPDFMDKPD------KTSYISERVIGKLFRKVK 936

Qy       1176 DKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEV- 1234

Db        937 DKA---------------------PQASSIATFTRDVARRSYDADMEVDGFEDYIDEAF 974

Qy       1235 --EKEYGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDKSGANYD--------SKV 1284

Db        975 DYKTEY-----DNKLGNLMDYYGIKTE------AEILSGGIMKASKTFDRRKDAEAISVA 1023

Qy       1285 IRLLELSFLADREMNTW---------ALLRASTAFKLYYHKS-------------PKFV 1321

Db       1024 VRALR------KEARAWFKRRNDIDDMLPKASAWYHVTYHPTYWGCYNQGLKRAHFISFP 1077

Qy       1322 WQMAGRQLAYIKAQMTSRP 1340

Db       1078 WCVYDQLIQIKKDKARNRP 1096
```

RESULT 3
T20695
hypothetical protein F10B5.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T20695; T24512
R;Sims, M.
submitted to the EMBL Data Library, February 1995
A;Reference number: Z19310
A;Accession: T20695
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1780 <WIL>
A;Cross-references: EMBL:Z48334; PIDN:CAA88315.1; GSPDB:GN00020; CESP:F10B5.7
A;Experimental source: clone F10B5
R;Burton, J.
submitted to the EMBL Data Library, October 1995

```
A;Reference number: Z19901
A;Accession: T24512
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1780 <WI2>
A;Cross-references: EMBL:Z66500; PIDN:CAA91312.1; GSPDB:GN00020; CESP:F10B5.7
A;Experimental source: clone T05C12
C;Genetics:
A;Gene: CESP:F10B5.7
A;Map position: 2
A;Introns: 23/2; 226/3; 305/2; 343/2; 462/3; 736/1; 922/2; 993/1; 1222/2; 1463/3; 1616/2

  Query Match             4.0%;  Score 296.5;  DB 2;  Length 1780;
  Best Local Similarity   20.4%;  Pred. No. 1.9e-09;
  Matches  196;  Conservative  129;  Mismatches  348;  Indels  289;  Gaps   37;

Qy        468 FPEKPPNDVFVTAMTGNFES---------KGSAVVLSAVLDYNPDNSPT------APLYL 512
Db        631 FDERRQIDV-VTAFTTMYQSRKIQYERLLKGESLQDVGLAKPLPKNCVSVAKVIVTPSRI 689

Qy        513 VKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQ 572
Db        690 LLMAPEVMMVN-RVVRRFGPD--------------------------------------- 709

Qy        573 HSLVGRQWRAFFAKDAGYRKPLREFQLRAED-PKPIIKERVHFFAETGITFRPDVFK--- 628
Db        710 YAL-----RCVFRDDNLGRLAIRDFSINNIDHMSNIVTEGIYLTLKNGIQVADRVYSFLG 764

Qy        629 --------------------TRSVVPAEEPV-----EQRTEFKVSQMLDWLLQLDNNTW 662
Db        765 WSNSQMRDQGCYLYAPRVNALTGEVTGTVEDIRVWMGDFRDAISVPKMMSRMGQCFTQA- 823

Qy        663 QPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMN-----DGVGRMSRS 717
Db        824 QP------TVYSSVKNIHIVENIQVRLERHHW--IVEPDIEGGVENKYCFSDGCGRISIK 875

Qy        718 VAKRIRDVLGLGDVPSAVQGRFGSAKGMWVID--VDDTGDEDWIETYPSQRKWECDFVDK 775
Db        876 LATHISKILQLKEVPACFQVRFKGFKGILVIDPTIDDIINMPKVIFRKSQQKFGEGGGEL 935

Qy        776 HQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQ--------------------- 814
Db        936 QDEYLEVVKYAMP-SPVCLNRPFITIL-----DQVSEKQSASSHRRITNRVHYYLERELC 989

Qy        815 AIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEET 874
Db        990 SLSNMLINENQ----AAEELVNRTNLAIDW--NAASKRA--------------------- 1022

Qy        875 LNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNI----RVGRSAYIYMIADFWGVL 930
Db       1023 -------GFELSVDPLIRDMLFSIYRYNIIHHISKAKIFLPPSLGRS--MYGVVDETGLL 1073

Qy        931 EENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVII 990
Db       1074 QYGQVFIQYSPSIR-QTSNRPILKTGKVLITKNPCHVPGDVRVFDAVWQPALAHLVDVVV 1132

Qy        991 FSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQL 1050
Db       1133 FPQHGPRPHPDEMAGSDLDGDEYSIIWDQEMLLDY-NEEAMVFP---------------- 1175

Qy       1051 MASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNS---VSNKPAIIL 1107
Db       1176 --SSSAAEEDKEPTTDDMVEFFLRY-LQQDSIGRMSH--AHLAYADLHGLFHENCHAIAL 1230

Qy       1108 SSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLK 1167
Db       1231 KCAVA--VDFPKSGVPAEPLS-------------SFEQCEMTPDYMMSGGKPMYYSTRLN 1275

Qy       1168 FSIARPA--IDKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFT 1225
Db       1276 GQLHRKARKVEEVLEEF-------------------------------ETRGSVF--- 1299

Qy       1226 TLKNRIGEVEKEYGRL-------------VKNKEMRDSKDPYPVRVNQVYEKWCAITPEA 1272
Db       1300 ---------EREYDKLICPEDVDVFFGNEIKLVQTLTLRDEYVDRMQQLLDEY-GIEDEA 1349

Qy       1273 MDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHK-----SPKFVWQMAGR 1327
Db       1350 SVVSG--------HAASIKRLAGMERDDYSFYHTDKVVELRYEKLYAVFRAKFFEEFGGE 1401

Qy       1328 QL 1329
Db       1402 EI 1403

RESULT 4
T01920
probable RNA-directed RNA polymerase (EC 2.7.7.48) - Arabidopsis thaliana
N;Alternate names: protein T22B4.110
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 10-Dec-1999
C;Accession: T01920; T08192
R;Strong, C.; Graves, T.; Duckels, G.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of A. thaliana F2P3.
A;Reference number: Z14455
A;Accession: T01920
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1133 <STR>
A;Cross-references: EMBL:AF080120; NID:g3600045; PID:g3600048
A;Experimental source: cultivar Columbia
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schue
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16098
A;Accession: T08192
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1133 <BEV>
A;Cross-references: EMBL:AL049876
A;Experimental source: cultivar Columbia; BAC clone T22B4
C;Genetics:
A;Map position: 4
A;Introns: 184/3; 820/2; 870/3
A;Note: F2P3.11; T22B4.110
C;Superfamily: Arabidopsis probable RNA-directed RNA polymerase
C;Keywords: nucleotidyltransferase

  Query Match             3.8%;  Score 279.5;  DB 2;  Length 1133;
  Best Local Similarity   21.8%;  Pred. No. 9.9e-09;
  Matches  156;  Conservative  103;  Mismatches  245;  Indels  211;  Gaps   31;

Qy        698 LSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVD------ 751
Db        544 VTTDGADYCFSDGIGKISLAFAKQVAQKCGLSHVPSAFQIRYGGYKGVIAVDRSSFRKLS 603

Qy        752 --------DTGDE-----DWIETYPSQRKWECDFVDKH----QRTLEVRSVASELKSAGL 794
Db        604 LRDSMLKFDSNNRMLNVTRWTESMP------C-FLNREIICLLSTLGIEDAMFEAMQA-V 655

Qy        795 NLQLL-PVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRA 853
Db        656 HLSMLGNMLEDR-----------DAALNVLQK---------------------------- 676

Qy        854 TRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIR 913
Db        677 ----------LSG--ENSKNLLVKMLLQGYAPSSEPYLSMMLRVHHESQLSELKSRCRIL 724

Qy        914 VGRSAYIYMIADFWGVLEENEVHVGFS-----------SKFR--DEEESFTLLSDCDVLV 960
Db        725 VPKGRILIGCMDEMGILEYGQVYVRVTLTKAELKSRDQSYFRKIDEETSVVI---GKVVV 781

Qy        961 ARSPAHFPSDIQRVRAVFKPELH----SLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVC 1016
Db        782 TKNPCLHPGDIRVLDAIY--EVHFEEKGYLDCIIFPQKGERPHPNECSGGDLDGDQFFVS 839
```

```
Qy       1017 WDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFA 1076
              || :|:   ::  |            |  :  |                   | : | | |
Db        840 WDEKIIPSEMDPPMDYAGSRPRLMDHD-------------------VTLEEIHKFFVDY 879

Qy       1077 LQPNFLGMCTNYKERLCYIN---NSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLR 1133
               :  : ||: :       | : :    ::  :  | ::|      || :|  |    | :|
Db        880 MISDTLGVIST--AHLVHADRDPEKARSQKCLELANLHSRAVDFAKTGAP-AEMPYALKP 936

Qy       1134 RELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLKFSIA--RPAIDKE------------ 1178
               ||          | | |:| |:      :    :| |:|  :|   : |
Db        937 REFPDFLERFEKPTYISESVFGK-----LYRAVKSSLAQRKPEAESEDTVAYDVTLEEAG 991

Qy       1179 LEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEY 1238
                |:|    || :|          :    | |  ::  :::   : :| |            ||
Db        992 FESFIETAKAHRD------MYGEKLTSLMIYYGAANEEEILTGILKT----------KEM 1035

Qy       1239 GRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPEAM---DKSGANYDSKVIRLLELSFLAD 1295
                 |:     | ||     : |  :::            |||    :||    :     |
Db       1036 YLARDNRRYGDMKDRITLSVKDLHK-------EAMGWFEKSCEDEQQK----------K 1077

Qy       1296 REMNTWALLRASTAFKLYY------HKSPK-----FVWQMAGRQLAYIKAQMTSR 1339
               :  : |            ||          |:  |      | | : |  |   |||:    |
Db       1078 KLASAW-----------YYVTYNPNHRDEKLTFLSFPW-IVGDVLLDIKAENAQR 1120
```

RESULT 5
T11660
probable RNA-directed RNA polymerase (EC 2.7.7.48) - fission yeast (Schizosaccharomyces
C;Species: Schizosaccharomyces pombe
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C;Accession: T11660
R;Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z17305
A;Accession: T11660
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1215 <BAR>
A;Cross-references: EMBL:Z98533; NID:e1071719; PID:e334112
C;Genetics:
A;Map position: 1L
A;Note: SPAC6F12.09
C;Keywords: nucleotidyltransferase

```
  Query Match             3.6%;  Score 265.5;  DB 2;  Length 1215;
  Best Local Similarity   20.7%;  Pred. No. 8e-08;
  Matches  207;  Conservative  147;  Mismatches  371;  Indels  275;  Gaps   42;

Qy        410 HEAPLAVAWEVTRLFMHCKV-----DLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFR 464
               | ||:            | |          : |||                | ||   |: |: |
Db        314 HNAPIG----------RCNVLVLSFSIRDES------------DKDDIAFLLHNLEKFN 350

Qy        465 GKPFPEKPPNDVFVTAMTGNFESKGS----------AVVLSAVLDYNPDNSPTAPLYLV 513
                          |    |  |  :  ::: : |            | :|  | |  |  :   | : |
Db        351 -----LKSQLDKVVFHLVPDYKHRCSLINDKEIEEEIAYLLQACLSKNLLSEIDLPIILA 405

Qy        514 KLKPLMFEQGCRLTRRFGPDRFFEILIPSPTS-TSPS--------------------- 549
                || |  |:            :|  ::: | |:  :||
Db        406 NLKKLSKERA---------KKFLRLILTSKTALINPSELDFTKSFVFYDLSSASSIHIKK 456

Qy        550 --VPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPI 607
                 | |                   | : : ||               :|| |   :| |
Db        457 LYVTPTT------------LRIVEDSL-----------EAGNR-VIRNF--------- 481

Qy        608 IKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLK 667
                |:  :  |      ||    : :|  :   ::   ::   :: |:|  : ::: |  : |
Db        482 -KDFANRFMRVQIT--DEYYKQKIRGGSDGFRNEKLYSRIQQLLTYGIKVGNQIYE--FL 536

Qy        668 LFSRIQLGLSKTYAIMT---LEPHQIRHHKTDLLSP--------------SGTGEV---- 706
                |   ||      |   :    |    |||    |                    | | |:
Db        537 AFGNSQLREHGAYFFASGSDLNAKQIREWMGDFSEINSVSKYAARMGQCFSTTKEINRFC 596

Qy        707 ---------------MNDGVGRMSRSVAKRIR-DVLGLGDVPSAVQGRFGSAKGMWVIDV 750
                             |||| | ||  :|:  :|   |||  |  |  |   ||  |: :
Db        597 VDISLQDDIVRNNHCFTDGVGMASLSVIRRLSLEVKNHDMFPSAFQFRMGGYKG--VLSL 654

Qy        751 DDTGDEDWIE---TYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRAR 807
                  ::  :     :|    |: :    |      |  ||||   :  |      :| ||:|:|: :||
Db        655 APPTKLEYHQGNLVFP--RRSQDKFKSFHS-TLEVIKI-SRFSNAHLNMQLITLLEGLGV 710

Qy        808 DKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFL-AG 866
               :|    :      :|   :      :|:||  |            | | |:|:       |: ||
Db        711 EKTVFLELTRSQLSKMNESINSKQKSIL----MLRDNVDEYHSTLII------ADFIQAG 760

Qy        867 LPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADF 926
                 :  :         |:|   ::                  | |:         :| |    :  |   :  ||:   :||
Db        761 FLERDDAFTENLLNLYYE-----------WVLR-----LIKEKQKVSVPKGAYLLGVADE 804

Qy        927 WGVLEEN----------------EVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSD 970
               | |: :                  :     | |      : ::   ::   |    :|||:|:   |  |
Db        805 TGTLKGHYDDAVLSVPEIFIQTDTSTSFGSYSTGKLKTRVIVGLC--IVARNPSLHPGD 862

Qy        971 IQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEM 1030
               :: :||    ||  ||:||:|   | | ||  :            |||  |||    |  ||    ::   ||
Db        863 VRVCKAVRCDELMHLKNVIVFPTTGDRSIPAMCSGGDLDGDEYTVIWDQRLLPKIVNYPP 922

Qy       1031 PLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKE 1090
                 ||     : :    |      | |:                    | ::: |      :: :  ||: :|  :
Db        923 LLESSPKKSI--DFLEGKPLI---------------DSVKEFFVNYIKYDSLGLISNAWK 965

Qy       1091 RLCYINNS----VSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDP 1146
                : :::      :      :  |:  :      || :|  |:       :      :|              ||
Db        966 AWAHDHDNNPEGIFGNVCLELAEMHSKAVDFAKSGVACKMQAKYHPKR--------YPDF 1017

Qy       1147 MYKSDSWLGRGEPTHIIDYLKFSI------ARPAIDKELEAFHNAMKAAKD--------- 1191
               | |:  :    |  | |   :       :::            |||            :  :      |
Db       1018 MQKTKTRSFRSE-TAVGKIFRYAARFQRESGRPA------TYNPIMNTVYDPCMKLPRFK 1070

Qy       1192 ------TEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNK 1245
                     |:      :|  |||   |            ::      :|      |    |: :  :       |||         :
Db       1071 TEYLNVAEEVKKHYDNDLRSIMARFDISTEYEVYTA--FILFKDDLAKTVNEYG---LRE 1125

Qy       1246 EMRDSKDPYPVRVNQVYEKWCAITPE-AMDKSGANYDSKV 1284
               |:    |     :   |  | :  ||:: :  |  |  |     |: ::
Db       1126 EVSFQFDLLKKKYTQEYLEKCALSNQSAFDSS--EYEERI 1163
```

RESULT 6
T21381
hypothetical protein F26A3.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T21381
R;McMurray, A.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19415
A;Accession: T21381
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3228 <WIL>
A;Cross-references: EMBL:Z78419; PIDN:CAB01701.1; GSPDB:GN00019; CESP:F26A3.3
A;Experimental source: clone F26A3
C;Genetics:
A;Gene: CESP:F26A3.3
A;Map position: 1
A;Introns: 36/3; 91/3; 216/3; 263/3; 308/3; 702/3; 828/3; 866/1; 1013/3; 1171/3; 1231/3
; 2443/1; 2595/3; 2663/1; 2753/3; 2813/3; 2859/1; 3052/3; 3136/1; 3176/1

```
  Query Match             3.1%;  Score 232.5;  DB 2;  Length 3228;
  Best Local Similarity   20.8%;  Pred. No. 4.3e-05;
  Matches  133;  Conservative  110;  Mismatches  217;  Indels  179;  Gaps   33;

Qy        701 SGTGEVMNDGVGRMSRSVAKRIRDVLGLGD-VPSAVQGRFGSAKGMWVIDVDDTGDE--D 757
                :|:     :||||  ||     |: :    :    |    |||    | ||    ||   ||  ::   ||
Db       2248 NGSEYTFSDGVGMMSYQFAQEVSQAMQFGKAVPSCFQIRFRGNKG--VIAIEPFLDEIRK 2305
```

```
Qy        758 WIET------------YPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLED 804
Db       2306 WALVNGVTSMKMAKCLFRPSQIKFQAKAISGDQ--IEMVKFSSAVLVA-LNKPFINIL-- 2360

Qy        805 RARDKVKMRQAIG------------DRLINDLQRQFSEQKHALNRPVEFRQWVYESYSS 851
Db       2361 ---DQVSEMQSLDCHKRITSRIEELMDRQILSFAKQMNEETFCRNKLKEFPR-------- 2409

Qy        852 RATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLN 911
Db       2410 ---RIDIDNLRTMWGFTLSSEPFFRSLIKA-----------SIKFSITKQLC---KEQIQ 2452

Qy        912 I--RVGRSAYIYMIADFWGVLEENEVHVGFSSKF------RDEEESF--TLLSDCDVLVA 961
Db       2453 IPSELGRS--MLGVVDETGRLQYGQIFVQYTKNYKKKLPPRDSNNKVHGSEIVTGTVLLT 2510

Qy        962 RSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEI 1021
Db       2511 KNPCIVPGDVRIFEAVDIPELHHMCDVVVFPQHGPRPHPDEMAGSDLDGDEYSVIWDQEL 2570

Qy       1022 VDGFVNAEMPLEP-DLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFA-LQP 1079
Db       2571 L-----LERNEEPFDFA--VEKIKVPY-------------DREKLDVLMREFYVTYLKL 2609

Qy       1080 NFLGMCTNYKERLCYINNS----VSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRE 1135
Db       2610 DSVGQISN-----SHLHNSDQYGLNSRVCMDLAKKNCQAVDFTKSG------------- 2650

Qy       1136 LLGGALSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDG 1195
Db       2651 ------QPPDPL--ETKW--RADP------VTFEVIPPENPERIPDFH------MGNERS 2688

Qy       1196 AHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYP 1255
Db       2689 PMYVSPRLCG--KLFRE-----------FQAIDN-----------VIKISEERD--EQYN 2722

Qy       1256 VRVNQV-----YEKWCAITPEAMDKSGANYDSKVIRLLE 1289
Db       2723 IELDETIFVTGFERYM----ESAQKQLSSYNGQLRSIME 2757
```

RESULT 7
T32650
hypothetical protein F39C12.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T32650
R;Chissoe, S.; Sansone, J.
submitted to the EMBL Data Library, December 1997
A;Description: The sequence of C. elegans cosmid F39C12.
A;Reference number: Z21206
A;Accession: T32650
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5105 <CHI>
A;Cross-references: EMBL:AF039043; PIDN:AAB94194.1; GSPDB:GN00028; CESP:F39C12.1
A;Experimental source: strain Bristol N2; clone F39C12
C;Genetics:
A;Gene: CESP:F39C12.1
A;Map position: X
A;Introns: 19/2; 47/3; 103/2; 142/2; 174/2; 324/2; 362/3; 436/3; 494/1; 523/1; 923/3; 98
88/3; 4971/3; 5020/3

```
  Query Match             2.6%;  Score 191;  DB 2;  Length 5105;
  Best Local Similarity   18.4%;  Pred. No. 0.032;
  Matches  303;  Conservative  207;  Mismatches  557;  Indels  578;  Gaps   81;

Qy         85 AASSNWVPKAHAD---PDTLPWSK--EPPRAATAGQQWALQTVLLEVLNR----FMPPPN 135
Db       2841 AESEEKVPVEKVDTPSPLTPPAASVPAPPPKMPADQQWRAAMMARRYMGASKPVPFVKPPS 2900

Qy        136 NTPGRTFGRTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLTR------------- 181
Db       2901 -----ALSKTESLKEASNIQESASTTRVQPALVVSHSTPEPVQLTRMAAAKKSEQKEYKT 2955

Qy        182 -----SATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQCTK--------RAKGKLSD 228
Db       2956 PVAKPKATAPP---------PITPKAASIRPATQKLHESELTKSEASRNVSSSSGKSQK 3005

Qy        229 NVAAAAAP---------PVP-IASALD-KVPTRRH---ANTRDPTATGHRRADQVDSFDT 274
Db       3006 QADVQSRPKLQSVQEKRSVPKVVKPLDPKGPQKEQESKKNEKVVPSTNVTLDDQASSSSN 3065

Qy        275 SQG--------------------------TSYGSSVFSACRHNQSTTQSSFEAPPSQPR 307
Db       3066 QQGPGSAKKPSASIFTALFGRKKEKKMTSSTAESSSIPIPSPENQEATSETNRSSPILPR 3125

Qy        308 EKRPVDATVFEAGHLIE-SPS---------KGRTTKSHIDNQPLSSSSQGETSFSTYYES 357
Db       3126 IKKP------ETGQLIKRSPTSDVPKPCVVSGAIVK-FVDGKPISNS------------ 3165

Qy        358 FPSSGGEGAIPEPSRSNGLA-----RSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEA 412
Db       3166 -PSYGKPMARRRPKKADSAADSGDLGSDEGSTS----------------------HEG 3200

Qy        413 PLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKP 472
Db       3201 --------------C---LKDEENSQK-KPAGKTV-----------RFGPISYEPPEPAP 3231

Qy        473 PNDVFVTAMTG----NFESKGSAVV--------LSAVLDYNPDNSPTAPLYLVKLKPLMF 520
Db       3232 PQPVEIPKVEPTPEISFESLAQQMAQKTLRKEEASKILEHYPVTEQTSPSTAEPVSPT-- 3289

Qy        521 EQGCRLTRRFGPDRFF----------------------EILIPSPTSTSPSVPPVVSKQ 557
Db       3290 KSGSKLS-FFGLSPFGKVTPKKNTAQTQSLTVLESTSSPELLTPSNPETSNTASTSSSKQ 3348

Qy        558 PGAVEE-------VIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREF----QLRAEDPKP 606
Db       3349 PKSLSEEKITATQMIQ--TRG-------------------RPIRKLSHRRQLKIGDPSP 3386

Qy        607 IIKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLD---------WLLQL 657
Db       3387 L--------QLSGDSSLPR--------PAQPSPEKPPQLEVPQGVQTYNPPSQRKYYTRL 3430

Qy        658 DNNTWQPHLK----LFSRI-QLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMN---- 708
Db       3431 -GREWKPPPKAKTPLSIRVPQSGVSTSTPAIDLSAKEV------LLAPIEKTKKMRKRKP 3483

Qy        709 ---------DGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGD---- 755
Db       3484 KPMETTLVLDGTEHFKKLTAENLQ--LLSRQFKSKRKNKKKRTKTWSVIGCDENDDSTKC 3541

Qy        756 ---EDWIETYPSQRK-----------------WECDFVD----KHQ----------RTLE 781
Db       3542 MQHNKLLEEFRDLKKSLITRQLDEKMITARAVFNPRFTDSSSTKHELLRNKPKFETTRKK 3601

Qy        782 VRSVAS----ELKSAGLNLQL---LPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHA 834
Db       3602 LKRLANNRQQKIRMSGIHSQPKRDVAISAIRQVQKVRFFQSTGKKMMNPIISGF-KPKSI 3660

Qy        835 LNRPVEFRQWVYESYSSRATRV-SHGRVPFLAGLPDS-----QEETLNFLMNSGFDPKKQ 888
Db       3661 LKPPKPPATWLF--------RIDSEGNKLYMKKLGPSDWQKKNDEKCNPM--SAMDPRKK 3710

Qy        889 K---YLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSSKFRD 945
Db       3711 KSNSYLK--RW-----------AKMMAKVGK--------------------------- 3728

Qy        946 EEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSG 1005
Db       3729 --KNMTKLFETNQLTRRLGAP-----------KKDSHP-KDVLVDPPKADAPPPSR--- 3770

Qy       1006 GDYDGDMAWVCWDP-----EIVDGFVNAEMPLEPDLSRYLKKD------KTTFKQLMASH 1054
Db       3771 -----SLSWTTVSSNASLNSLLDA-LNTVGSDEPSTSSGIVTDIAPPIVKEQFRKAKESR 3824
```

```
Qy       1055 GTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCY---------INNSVSNKPAI 1105
Db       3825 PTQQTVSKK--YPTVYLGMTVARSPSSSTSLVNLDEREAVSSLAQDGSEIVEYVGQEEPV 3882

Qy       1106 I-------------LSSLVGNLVDQSKQGIVFNEASWAQLRRELLG----GALSLPD--P 1146
Db       3883 VETTKRRFDPFVRPISALFKAFTSSSPEVGLEADSRSRPTERTLRGLFRRSSPTMPEFST 3942

Qy       1147 MYKSDS----WLGRGEPTHII------DYLKFSIARP----AIDKELEAFHNAMKAAKDT 1192
Db       3943 RHVSSSPTRRRANRGPVTSVSQGSSAEDILALMQAQPDALENMEEQVVSKFPPIQLQAPV 4002

Qy       1193 EDGAHFWDPDLA---SYYTFFKEISDKSRSSALL------------FTTLKNRIGEVEKE 1237
Db       4003 LNGKSMWRPKLKLKNSLYTQFKELKEESKKRKLVEYLDQNYKNNKSYSTLQTLITFNLDQ 4062

Qy       1238 YGRL------VKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDKSGANYDSKVIRLLELS 1291
Db       4063 YMRTNLYPRPKLNTDLLGKYDPNSLRVS------------SSSRSGLQRSTRLI--VKKI 4108

Qy       1292 FLADREMNTW-ALLRASTAFKLYYHKSPKFVWQMAGRQLAYIKAQMTSRPGEGAPALM-- 1348
Db       4109 RRCQEQLEKWKRLIATSTNQQL-----------TNGELVKRAKASMKQASNRKAKKVLVL 4157

Qy       1349 -----TAFMY----------AGLMPDKKFTKQYVARLEGD--------GSEYPDPEVYEV 1385
Db       4158 YKFKRTCFKYRVQVTITKSARKILADQNQKKQLIMHEEKKRTPTILTIGDEDDENEIVDL 4217

Qy       1386 L----------------GDDDFDGI 1394
Db       4218 LSGNEDSANCQNLADYDGDNDDDAI 4242
```

RESULT 8
T38253
hypothetical protein SPAC23C11.15 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T38253
R;Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1995
A;Reference number: Z21781
A;Accession: T38253
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1075 <BRO>
A;Cross-references: EMBL:Z98559; PIDN:CAB11171.1; GSPDB:GN00066; SPDB:SPAC23C11.15
A;Experimental source: strain 972h-; cosmid c23C11
C;Genetics:
A;Gene: SPDB:SPAC23C11.15
A;Map position: 1
A;Introns: 11/1; 71/2

```
  Query Match             2.5%;  Score 182;  DB 2;  Length 1075;
  Best Local Similarity   18.7%;  Pred. No. 0.0086;
  Matches 193;  Conservative 157;  Mismatches 359;  Indels 322;  Gaps   51;

Qy        494 LSAVLDYNPDNSPTAPLYL-VKLKPLMFEQGCRLTRRFGPDRFFEILIPS-PTSTSPSVP 551
Db         81 ISVILRDYPDLLEYLNIFLPSSYKYLLSNSGANFTLQF--------TTPSGPVSTPSTYV 132

Qy        552 PVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQ------------- 598
Db        133 ATYNDLPCTYHRAIGFVSRVRRALLSNPEQFFKLQDS-----LRKFKNSECSLSELQTIV 187

Qy        599 --LRAEDPKPIIKERVHFFAETGITF--RPDV--FKTRSVVPAEEPVEQRTEFKVSQMLD 652
Db        188 TSLLAEHPS--LAHEFHNFLPSSIFFGSKPPLGSFPLRGI--------QSSQFTLSNISD 237

Qy        653 WLLQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVG 712
Db        238 LLSQ----------------------------------SRPDNLSP--FSHLSNE--- 256

Qy        713 RMSRSVAKRIRDVL------------------GLGDVPSAVQGRFG---SAKGMW--VID 749
Db        257 --SSDFFKNVKNVLTDVETYHEFLKLLNLYVQGIIDRNILVSRGFGFLKSNSGLWRSFLS 314

Qy        750 VDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDK 809
Db        315 LTSLSPEEFLSVY-------------------NSACSDFPECGPSYRLLPVEERNISCS 354

Qy        810 VKMRQAIGDRLIND----------------LQR---------QFSEQKHALNRPVEFRQW 844
Db        355 GRDDFAWG--ILNDDWVSHPTWASEESGFIVQRKTPYEEAMTKLEEERYEFDRHIEATSW 412

Qy        845 VYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKK-----------QKYLQD 893
Db        413 TIKSLKKIQNRINE--------LPEEERETYTLEEGLGLPSKSIYKKTIKLVYTSEHAEE 464

Qy        894 IAWDLQKRKCDTLK---SKLNIR------VGRSAYIYMIADFWGVLEENEVHVGFSS--- 941
Db        465 MFKALERMPCLTLPLVISRLEEKNEEWKSVKRS-----LQPGWRSIEFKNYDKSLDSQCV 519

Qy        942 --KFRDEE--ESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDV 997
Db        520 YFKARDKKNVSSKFLLAEADILRSQAKLHFP---LRSRSAFEFSFVYDNEIVLFDT---- 572

Qy        998 PLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTG 1057
Db        573 ------------------CY---MVCTYIVCNSP------SGLKKVEHFFKNILPLHFGL 605

Qy       1058 SAAKEQTTYDMIQKSFHFALQ-PNFLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVD 1116
Db        606 EKDKFSIFLDQVFRGPDYDVNAPNIVG------------NKPVRRKRSNSITQLT-EFVK 652

Qy       1117 QSK-QGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAI 1175
Db        653 QPKINGQRESRSAAAARKKEESGNKSQSNSQNSLSD------ESGNVTPVSKKQLSQPAA 706

Qy       1176 DKELEAFHNAMKA-----------------AKDTEDGAHFWDPD-----LASYYTFFKEI 1213
Db        707 ---------AIKASLKYPSHPDSLLEHQDHAGDTENEMHD-DVDKEQFGYSSMYVFFR-- 754

Qy       1214 SDKSRSSALLFTTLKNRIGEVEK--EYGRLVKNKEMRDSKDPYPVRVNQVY--EKWCAIT 1269
Db        755 ---------LFNLLYERLYELQRLEDQVSIIQQRII-----PNPVSQKQKIWRDRWNDLS 800

Qy       1270 PEAMDKSGANYDSKVIRLLELSF-LADREM--NTWALLRASTAFKLYYHKSPKFVWQMAG 1326
Db        801 DVPDEKT--HYENTYVMILRLIYGIVDQSAFEDYLRFYYGNKAYKIY--TIDKLVWS-AA 855

Qy       1327 RQLAYIKAQMTSRPGEGAPALMTAFM--YAGLMPDKKFTKQYVARLEGDGSEYPDPEVYE 1384
Db        856 KQVHHIVS-------DGKYKFVTSLVEQNSSASPKKNY-DDFLYRLEIEKLLNPDEILFR 907

Qy       1385 VLGDDDFDGIG 1395
Db        908 FCWINKFKSFG 918
```

RESULT 9
F84582
hypothetical protein At2g19910 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: F84582
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, I
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: F84582
A;Status: preliminary
A;Molecule type: DNA

```
A;Residues: 1-966 <STO>
A;Cross-references: GB:AE002093; NID:g3687227; PIDN:AAC62125.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g19910
A;Map position: 2

  Query Match             2.4%;  Score 177.5;  DB 2;  Length 966;
  Best Local Similarity   19.8%;  Pred. No. 0.014;
  Matches  167;  Conservative  116;  Mismatches  267;  Indels  293;  Gaps   41;

Qy        518 LMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVG 577
Db        232 LLENTGTHLHKVLGDDNVLTVKFDKVLGVETYCNDLYSTYKGIAKNGI---------MVG 282

Qy        578 -RQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVH-FFAETGITF-----RPDVFKTR 630
Db        283 LRRYRFFVFKDGG----------KEEKKKDVSTKGVKCYFIRTDSTASIDMQNPYIFAGK 332

Qy        631 SVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQI 690
Db        333 SM------HEARMHFM------------------HVNTLSSLPNYMSR---------YSI 359

Qy        691 RHHKTDLLSPSGTGEVMNDGVGRMSRSVA----------KRIR---------DVLGLG-- 729
Db        360 DQDDKDVLDKNGKPCIHSDGTGYISEDLARMCPVNIFKGKSMRSNNIQSKNLNFEGQGPC 419

Qy        730 --DVPSAVQGRF----GSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVR 783
Db        420 GQEPPLLIQFRIFYNGYAVKG----------------TFLTNKKLP-------PRTVQVR 456

Qy        784 SVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQ---RQFSEQKHALNRPVE 840
Db        457 P------------SMIKVYEDRTLSNLS--------TFNSLEVVTTSNPPRKARLSR--- 493

Qy        841 FRQWVYESYSSRATRVSHGRVP---FLAGLPDSQEETLNFL----------MNSGFD--- 884
Db        494 ----------NLVALLSYGGVPNDFFLNILRNTLEESKTIFYSERAAFKAAINYGDDQYT 543

Qy        885 --------PKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEV- 935
Db        544 ADMILVGIPLDEPYLKDRLSYLLKTERNALKAG-RFPIDESYYIMGTVDPTGELKENEIC 602

Qy        936 ---HVGFSSKFRDEEESFTLLSDCDVLVARSPA-HFPSDIQRVRAVFKPELH----SLKD 987
Db        603 VILHSGQISG--------------DVLVYRNPGLHF-GDIHVLKATYVKALEDYVGNAKF 647

Qy        988 VIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTF 1047
Db        648 AVFFPQKGPRSLGDEIAGGDFDGDMYFISRNPKLLEHF----KPSEPWVSS--SKPSKIY 701

Qy       1048 KQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNSVSNKPAIIL 1107
Db        702 --------CGRKPSELSEEELEEELFKMFL-----------KARFC-------------- 728

Qy       1108 SSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHII--DY 1165
Db        729 -------------------------KRDVIGMA---------ADCWLGIMDPFLTLGDES 754

Qy       1166 LKFSIARPA-IDKELEAFHNAMKAAKDTEDGAHF-WDPDL-ASYYTFFKEISDKS--RSS 1220
Db        755 AKEKYERKKNILKLIDIYYDALDAPK---KGAKVDLPPDLEIKNFPHYMERDPKRDFRST 811

Qy       1221 ALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPEA-MDKSGAN 1279
Db        812 SIL---------------GLIFDTVDSHNAEEPPPSEISKLWYFEDEPVPKSHMDKFTSW 856

Qy       1280 YDS 1282
Db        857 YEN 859

RESULT 10
S15053
hypothetical protein YCR033w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YCR592
C;Species: Saccharomyces cerevisiae
C;Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 19-Apr-2002
C;Accession: S15053; S19445; S40918
R;Jia, Y.; Slonimski, P.P.; Herbert, C.J.
Yeast 7, 413-424, 1991
A;Title: The complete sequence of the unit YCR59, situated between CRY1 and MAT, revea
A;Reference number: S15052; MUID:91335897; PMID:1872032
A;Accession: S15053
A;Molecule type: DNA
A;Residues: 1-1226 <JIA>
A;Cross-references: EMBL:X59075; NID:g4805; PIDN:CAA41799.1; PID:g4807
R;Herbert, C.J.; Jia, Y.; Slonimski, P.P.
submitted to the Protein Sequence Database, March 1992
A;Reference number: S19445
A;Accession: S19445
A;Molecule type: DNA
A;Residues: 1-1226 <HER>
A;Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42300.1; PID:e264374; PID:g1907
R;Wickstead, B.L.; Roberts, A.B.; Sagliocco, F.A.; Brown, A.J.P.
Yeast 7, 761-772, 1991
A;Title: The complete sequence of a 7.5 kb region of chromosome III from Saccharomyces
A;Reference number: S40917; MUID:92133166; PMID:1776366
A;Accession: S40918
A;Molecule type: DNA
A;Residues: 1-304,'A',306-374,'S',376-403,'Q',405-434,'V',436-441,'N',443-481,'E',483-
>
A;Cross-references: EMBL:S78624; NID:g244237; PIDN:AAB21259.1; PID:g244239
C;Genetics:
A;Cross-references: SGD:S0000629
A;Map position: 3R

  Query Match             2.4%;  Score 174;  DB 2;  Length 1226;
  Best Local Similarity   18.4%;  Pred. No. 0.033;
  Matches  218;  Conservative  178;  Mismatches  453;  Indels  338;  Gaps   52;

Qy         84 KAASSNWVPKAHADPDTLPWSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFG 143
Db         38 KSSNNGFVSSPTADNSTNP-SVTPSTAS-------------------VPLPTAAPGSTFG 77

Qy        144 -----RTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFP 198
Db         78 IEAPRPSRYDPSSVSRPSSSSYSSTRKIGSRYNPDVERSSSTTSSTPESMNTSTITHTNT 137

Qy        199 DPVNTGSKRPSLESENLNQCTKRAKGKLSDNV-----AAAAAPPVPIASALDKVPTRRHA 253
Db        138 DIGNSRYSRKTMSRYN-------PQSTSSTNVTHFPSALSNAPPFYVANGSSRRP--RSM 188

Qy        254 NTRDPTATGHRRADQVDSFDTSQGTSYGS------SVFSACR-------HNQSTTQSSFE 300
Db        189 DDYSPDVTNKLETNNVSSVNNNSPHSYYSRSNKWRSIGTPSRPPFDNHVGNMTTTSNTNS 248

Qy        301 APPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQ--------PLSSSSQGETSFS 352
Db        249 IHQREPFWKAN-STTILKSTHSQSSPSL-HTKKFHDANKLDKPEASVKVETPSKDETKTI 306

Qy        353 TYYE-SFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWL-- 409
Db        307 SYHDNNFPPRKSVSKPNAPLEPDNIKVGEEDALGKKEVHKSG-----REIAKEHPTPVKM 361

Qy        410 --HEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDV---TDIWKTLYRLDAFR 464
Db        362 KEHDELEARAKKVNKI-----------NIDGKQDEIWTTAKTVASAVEVSKESHK----- 405

Qy        465 GKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLV-------KLKP 517
Db        406 ---------------ELTRSVERKESPEIRDYERAYDPKALKTDATKLTVDDDNKSYEEP 450

Qy        518 LMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQ--HSL 575
Db        451 LEKVEGC--------------IFPLPKA-----------------ETRLWELKNQKRNKI 479
```

```
Qy        576 VGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPA 635
Db        480 ISKQKYLL-------KKAIRNF---SEYPFYAQNKLIHQQA-TGLI----LTKIISKIKK 524

Qy        636 EEPVEQRTEFKVSQMLDWL-LQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHK 694
Db        525 EEHLK-----KINLKHDYFDLQ---KKYEKECEILTKLSENLRKEEIENKRKEHELMEQK 576

Qy        695 TDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTG 754
Db        577 R-----------REEGI---ETEKEKSLRH-------PSS-------------------- 595

Qy        755 DEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRAR------- 807
Db        596 ------SSSSRRRNRADFVDDAEMENVLLQIDPNYKHYQAAATIPPLILDPIRKYSYKFC 649

Qy        808 ---DKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFL 864
Db        650 DVNNLVTDKKLWASRILKDASDNFTDHEHSL-----FLEG-YLIHPKKFGKISH----YM 699

Qy        865 AGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIA 924
Db        700 GGLRSPEECVLHY-----YRTKKTVNYKQLLIDKNKKRKMSAAAKRRKRKERS------- 747

Qy        925 DFWGVLEENEVHVGFSSKFR----DEEESF----------TLLSDCDV------------ 958
Db        748 ------NDEEVEVDESKEESTNTIDKEEKSENNAEENVQPVLVQGSEVKGDPLGTPEKVE 801

Qy        959 -LVARSPAHFPSDIQRVRAVFKPELHSLKDVI--------IFSTKGDVPLAKKLSGGDYD 1009
Db        802 NMIEKRGEEFAGELENAERV--NDLKRAHDEIGEESNKSSVIETNNEVQIMAP-KGGVRN 858

Qy       1010 GDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMI 1069
Db        859 G------YYPEETK---ELDFSLENALQR--KKHK-------------SAPEHKTSYWSV 894

Qy       1070 QKSFHF-----------ALQPNFLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQS 1118
Db        895 RESQLFPELLKEFGSQWSLISEKLGTKSTTMVRNYYQRNAARNGWKLLVDE-----TDLK 949

Qy       1119 KQGIVFNEASWAQL----RRELLGGALSLPDPMYKSDSWLGRGEPTH 1161
Db        950 RDGTSSESVQQSQILIQPERPNINAYSNIP-PQQRPALGYFVGQPTH 995
```

RESULT 11
G84582
hypothetical protein At2g19920 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: G84582
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G84582
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-905 <STO>
A;Cross-references: GB:AE002093; NID:g3687226; PIDN:AAC62124.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g19920
A;Map position: 2

```
  Query Match             2.3%;  Score 169.5;  DB 2;  Length 905;
  Best Local Similarity   21.6%;  Pred. No. 0.038;
  Matches  134;  Conservative  100;  Mismatches  209;  Indels  178;  Gaps   34;

Qy        575 LVG-RQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVH-FFAETGITFRPDVFKTRSV 632
Db        285 MVGLRRYQFFVFKDGG----------KAEKKKRNSTKQVKCYFIRTGSTASSDM------ 328

Qy        633 VPAEEPV--------EQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKT----- 679
Db        329 ---ENPYILSGMSIHEARMHF-----------MHVHTLPSPANYMARFSLILSKTKKLEV 374

Qy        680 --YAIMTLEPHQIRHH---KTDLLSPSGTGEVMNDGVGRMSRSVAK------------RI 722
Db        375 DMTEITVMQIDDIHCHDQSNNDVLDKNGKPRIHSDGTGYISEDLARMCPLNIFKGKSMRS 434

Qy        723 RDVLG--LGDVPSAVQGRF---GSA-KGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKH 776
Db        435 NNIQGTCVQEPPLLIQIRMFNDGSAVKGIFLL--NKNLPPQTVQVRPSMIK---VYKDKN 489

Qy        777 Q------RTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIG--------DRLIN 822
Db        490 LSNFSTFNSLEVVTTSNPPKRAKLSKNLVALL------------SYGGVPNDFFLDILLN 537

Qy        823 DLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSG 882
Db        538 TLEK---------------KKTIF--FKVRAA----GKAALHYGNMD-DKNALQMIM-AG 574

Qy        883 FDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSSK 942
Db        575 I-PLDEPYLKHYLSKLLKLEKDDLKAG-KLPIDESYYLMGTVDPTGELKEDEV------- 625

Qy        943 FRDEEESFTLLSDCDVLVARSPA-HFPSDIQRVRAVFKPELH----SLKDVIIFSTKGDV 997
Db        626 -------------CVIL---NPGLHF-GDIHILKATYVKSLEQYVGNSKYGVFFPQKGPR 668

Qy        998 PLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTG 1057
Db        669 SLGDEIAGGDFDGDMYFISRNPKLLEHY----KPSEPWVSSSPRSKIYTGRQ--PSELSP 722

Qy       1058 SAAKEQTTYDMIQKSFH------------FALQPNFLGM-CTNYKERLCYINNSVSNKPA 1104
Db        723 EQLEEELFKIFLKTGFSPSSVIGQAADSWLAIMDRFLTLGDENVKEK------AEMKKKM 776

Qy       1105 IILSSLVGNLVDQSKQGIVFN 1125
Db        777 LKLTDIYYDAIDAPKTGTEVN 797
```

RESULT 12
A55575
ankyrin 3, long splice form - human
N;Alternate names: ankyrin G
C;Species: Homo sapiens (man)
C;Date: 03-Mar-1995 #sequence_revision 03-Mar-1995 #text_change 20-Sep-1999
C;Accession: A55575
R;Kordeli, E.; Lambert, S.; Bennett, V.
J. Biol. Chem. 270, 2352-2359, 1995
A;Title: Ankyrin-G. A new ankyrin gene with neural-specific isoforms localized at the a
A;Reference number: A55575; MUID:95138209; PMID:7836469
A;Accession: A55575
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-4377 <KOR>
A;Cross-references: GB:U13616; NID:g608024; PIDN:AAA64834.1; PID:g608025
C;Genetics:
A;Gene: GDB:ANK3
A;Cross-references: GDB:424503; OMIM:600465
A;Map position: 10q21-10q21
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homolog
C;Keywords: alternative splicing; peripheral membrane protein
F;73-105/Domain: ankyrin repeat homology <AN01>
F;106-138/Domain: ankyrin repeat homology <AN02>
F;139-171/Domain: ankyrin repeat homology <AN03>
F;172-200/Domain: ankyrin repeat homology <AN04>
F;201-233/Domain: ankyrin repeat homology <AN05>
F;234-266/Domain: ankyrin repeat homology <AN06>
F;267-299/Domain: ankyrin repeat homology <AN07>
F;300-332/Domain: ankyrin repeat homology <AN08>
F;333-365/Domain: ankyrin repeat homology <AN09>

```
F;366-398/Domain: ankyrin repeat homology <AN10>
F;399-431/Domain: ankyrin repeat homology <AN11>
F;432-464/Domain: ankyrin repeat homology <AN12>
F;465-497/Domain: ankyrin repeat homology <AN13>
F;498-530/Domain: ankyrin repeat homology <AN14>
F;531-563/Domain: ankyrin repeat homology <AN15>
F;564-596/Domain: ankyrin repeat homology <AN16>
F;597-629/Domain: ankyrin repeat homology <AN17>
F;630-662/Domain: ankyrin repeat homology <AN18>
F;663-695/Domain: ankyrin repeat homology <AN19>
F;696-728/Domain: ankyrin repeat homology <AN20>
F;729-761/Domain: ankyrin repeat homology <AN21>
F;762-794/Domain: ankyrin repeat homology <AN22>
F;795-827/Domain: ankyrin repeat homology <AN23>

  Query Match             2.2%;  Score 163.5;  DB 2;  Length 4377;
  Best Local Similarity   17.3%;  Pred. No. 1.2;
  Matches  275;  Conservative  205;  Mismatches  538;  Indels  575;  Gaps   72;

Qy         30 CVADTTLTPHRRKELAESDEDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAASSN 89
Db       1430 CNLNITLPAHKKETESDQDDEIEKTDRRQSFASLALRKRYSYLTEPG---MIERSTGATR 1486

Qy         90 WVPKAHA---------------DPDTLPW------------SKEPPRAATAGQQWALQTV 122
Db       1487 SLPTTYSYKPFFSTRPYQSWTTAPITVPGPAKSGFTSLSSSSSNTPSASPLKSIWSVST- 1545

Qy        123 LLEVLNRFMPPP-NNTPGRTFGRTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLTR 181
Db       1546 ---------PSPIKSTLGASTTSSVKSISDVASPI-RSLRTMSSPIKTVVSQSPYNIQVS 1595

Qy        182 SAT---------GPPIHGAAIPLKF---PDPVNTGSK----------RPSLESENLNQCT 219
Db       1596 SGTLARAPAVTEATPLKGLASNSTFSSRTSPVTTAGSLLERSSITMTPPASPKSNINMYS 1655

Qy        220 KRAKGKLSDNVAAAAAPPV---------PIASALD-----KVPTRRHANTRDPTATGHRR 265
Db       1656 SSLPFK---SIITSAAPLISSPLKSVVSPVKSRVDVISSAKITMASSLSSPVKQMPGHAE 1712

Qy        266 ADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIES 325
Db       1713 VALVNGSISPLKYASSSTLINGCKATATLQEKISSATNS-------VSSVVSAATDTVEK 1765

Qy        326 PSKGRTT------KSHIDNQPLSSSSQGETSFSTYYESFPSS------------------ 361
Db       1766 VFSTTTAMPFSPLRSYVSAAPSAFQSLRTPSASALYTSLGSSISATTSSVTSSIITVPVY 1825

Qy        362 GGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEVT 421
Db       1826 SVVNVLPEPALKK-LPDSNSFTKSAAALLSPIKTLTTET---------HPQP---HFSRT 1872

Qy        422 RLFMHCKVDLEDESLGLKYDPSWSTA----RDVTDIWKTLYRLDAF------RGKPFPEK 471
Db       1873 SSPVKSSLFLAPSALKLSTPSSLSSSQEILKDVAEMKEDLMRMTAILQTDVPEEKPFQPE 1932

Qy        472 PPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTR--- 528
Db       1933 LPKE-------GRIDDEEPFKIVEKVKE-----------DLVKVSEILKKDVCVDNKGSP 1974

Qy        529 ----------------RFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEE-------VI 565
Db       1975 KSPKSDKGHSPEDDWIEFSSEEIREARQQAAASQSPSLPERVQVKAKAASEKDYNLTKVI 2034

Qy        566 QWLT--MGQHSLVGRQWRAFFAKDAG------YRKP---LREFQLR--------AEDPKP 606
Db       2035 DYLTNDIGSSSLTNLKYKFEDAKKDGEGGQKRVLKPAIALQEHKLKMPPASMRTSTSEKE 2094

Qy        607 IIKERVHFFAETGITFRPD---------------VFKTRSVVPAEEPVEQRTEFKVSQML 651
Db       2095 LCKMADSFFGTDTILESPDDFSQHDQDKSPLSDSGFETRSEKTPSAP--QSAE------- 2145

Qy        652 DWLLQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLL------SPSGTGE 705

Db       2146 ---------TTGPK-PLFHEVPI------------PPVITETRTEVVHVIRSYDPSA--- 2180

Qy        706 VMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQ 765
Db       2181 ----------------------GDVP-------------------QTQPEEPVSPKPSP 2198

Qy        766 RKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQ 825
Db       2199 TFME----------LEPKPTTSSIK-----------------EKVKAFQ---------MK 2222

Qy        826 RQFSEQKH--ALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNF-----L 878
Db       2223 ASSEEDDHNRVLSKGMRVKE---ETHITTTTRMVYHSPPGGEGASERIEETMSVHDIMKA 2279

Qy        879 MNSGFDPKKQ-----KYLQDIAWDLQKRKCDT----------LKSKLNIRVGRSAYIYMI 923
Db       2280 FQSGRDPSKELAGLFEHKSAVSPDVHKSAAETSAQHAEKDNQMKPKLERII--------- 2330

Qy        924 ADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELH 983
Db       2331 ----------EVHI----------EKGNQAEPTEVIIRETKKH-------------PE-- 2355

Qy        984 SLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKD 1043
Db       2356 --KEMYVYQ--------KDLSRGDI-------------------------NLKDFLPEK 2379

Qy       1044 KTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNSVSNKP 1103
Db       2380 HDAFP---CSEEQGQQEEEELTAE-------------------------------ESLP 2404

Qy       1104 AIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHII 1163
Db       2405 SYLESSRVNTPVSQEED----SRPSSAQL-----------ISDDSYKTLKLLSQ----HSI 2446

Qy       1164 DY-----------LKFSIARPAIDKELEAFHNAMKAAKDTEDGA--HFWDPDLASYYTF 1209
Db       2447 EYHDDELSELRGESYRFA-EKMLLSEKLDVSHS------DTEESVTDHAGPPS------- 2492

Qy       1210 FKEI--SDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQV------ 1261
Db       2493 -SELQGSDK-RSREKIATAPKKEI--LSKIYKDVSENGVGKVSKDEHFDKVTVLHYSGNV 2548

Qy       1262 ----YEKWCAITPEAMDKSGAN--YDSKVIR--------LLELS-FLADR------EMNT 1300
Db       2549 SSPKHAMWMRFTEDRLDRGREKLIYEDRVDRTVKEAEEKLTEVSQFFRDKTEKLNDELQS 2608

Qy       1301 WALLRASTAFKLYYHKSPK-------FVWQMAGRQLAYIKAQMTSRPGEG-------APA 1346
Db       2609 PEKKARPKNGKEYSSQSPTSSSPEKVLLTELLASNDEWVKARQHGPDGQGFPKAEEKAPS 2668

Qy       1347 LMTAFMYAGLMPDKKFTKQYV----ARLEGDGS 1375
Db       2669 LPSS-------PEKMVLSQQTEDSKSTVEAKGS 2694
```

RESULT 13
T30868
RhoA-binding protein p116Rip - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T30868
R;Gebbink, M.F.B.G.; Poland, M.; Kranenburg, O.; VanHorck, F.P.G.; Moolenaar, W.H.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z20911
A;Accession: T30868
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1024 <GEB>
A;Cross-references: EMBL:U73200; NID:g1657836; PID:g1657837; PIDN:AAB18198.1
A;Experimental source: brain
C;Function:
A;Description: may inhibit RhoA-stimulated contractility and promote neurite outgrowth

```
  Query Match             2.2%;  Score 160.5;  DB 2;  Length 1024;
  Best Local Similarity   19.9%;  Pred. No. 0.17;
  Matches  211;  Conservative  153;  Mismatches  407;  Indels  291;  Gaps   51;

Qy        161 KRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQCTK 220
Db        159 KRKVEPPTPQEPGPAKMAVTSSSGGTSGSSSSIP--------SAEKVPTTKS-TLWQEEM 209

Qy        221 RAKGKLSDNVAAAAAPP---VPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQG 277
Db        210 RAKDQPDGTSLSPAQSPSQSQPPAACTPREP-GLESKEDESTISG-------DRVDGGRK 261

Qy        278 TSYGSSVFSACRHNQSTTQSSFEAP------PSQPREKRPVDATVFEAGHLIESPSKGRT 331
Db        262 VRVESGYFSLEKAKQDLRAEEQLPPLLSPPSPSTPHSRRSQVIEKFEA--LDIEKAEHME 319

Qy        332 TKSHIDNQPLSSSSQGET---------SFSTYYESFPSSGGEGAIPEPSRSNGLARSEES 382
Db        320 TNMLILTTPSSDTRQGRSERRAIPRKRDFASEAPTAPLSDACPLSPHRRAKSLDRRSTES 379

Qy        383 ARSQVQVHAPVVAARLRNIWPKFPK-WLHEAPLAVAWE------VTRLFMHCKVDLEDES 435
Db        380 S----------MTPDLLN----FKKGWLTKQYEDGQWKKHWFVLADQSLRYYRDSVAEEA 425

Qy        436 LGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLS 495
Db        426 ADLDGEINLSTCYDVTE-----YPVQRNYGFQIHTK-EGEFTLSAMTSGIRRNWIQTIMK 479

Qy        496 AVLDYNPDNSP--TAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPV 553
Db        480 HVL---PASAPDVTSSLPEGKNKSTSFETCSRSTEK-------------QEAEPGEPDP 522

Qy        554 VSKQPGAVE-------EVIQWLTMG--QHSLVGRQWRAFFAKDAGYRKPL--------RE 596
Db        523 EQKKSRARERRREGRSKTFDWAEFRPIQQALAQERASAVGSSDSGDPGCLEAEPGELERE 582

Qy        597 FQLRAEDPK-----------PIIKE---RVHFFAETGITFRPDVFKTRSV-VPAEE---- 637
Db        583 RARRREEPRKRFGMLDTIDGPGMEDTALRMDIDRSPGLLGTPDL-KTQNVHVEIEQRWHQ 641

Qy        638 ----PVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQ-LGLSKTYAIMTLEPHQIRH 692
Db        642 VETTPLREEKQVPIAPL---------------HLSLEDRSERLSTHELTSLLEKELEQSQK 687

Qy        693 HKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVD- 751
Db        688 EASDLL--------------EQNRLLQDQLRVAL----------GREQSAREGYVLQATC 723

Qy        752 DTGDEDWIETYPSQRKWECDFVDKHQRTLE-VRSVASEL--KSAGLNLQLLPVLEDRARD 808
Db        724 ERGFAAMEETH--QKKIE-DLQRQHQRELEKLREEKDRLLAEETAATISAIEAMKNAHRE 780

Qy        809 KVKMRQAIGDRL--------INDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGR 860
Db        781 EMERELEKSQRSQISSINSDIEALRRQYLEELQSVQRELEV---LSEQYSQKCLENAH-- 835

Qy        861 VPFLAGLPDSQEETLNFLM--NSGFDPKKQKYLQDIAWDLQKRKCDTL--------KSKL 910
Db        836 ---LAQALEAERQALRQCQRENQELNAHNQELNNRLAAEITRLR--TLLTGDGGGESTGL 890

Qy        911 NIRVGRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSD 970
Db        891 PLTQGKDAY-----------------------------------ELEVLLRVK----ESE 911

Qy        971 IQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEM 1030
Db        912 IQ----YLKQEISSLKDELQTALRD-----KKYASDKY----------KDIYTELSIAKA 952

Qy       1031 PLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTT---YDMIQKSFHFALQPNFLGMCTN 1087
Db        953 KADCDISRLKEQLKA------ATEALGEKSPEGTTVSGYDIMKSKSN----PDFL----- 997

Qy       1088 YKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASW 1129

Db        998 KKDRSC-------------VTRQLRNIRSKS----VIEQVSW 1022
```

RESULT 14
JN0583
myosin-light-chain kinase (EC 2.7.1.117) 155K protein - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 20-Apr-2000
C;Accession: JN0583; PN0490
R;Kobayashi, H.; Inoue, A.; Mikawa, T.; Kuwayama, H.; Hotta, Y.; Masaki, T.; Ebashi, S
J. Biochem. 112, 786-791, 1992
A;Title: Isolation of cDNA for bovine stomach 155kDa protein exhibiting myosin light ch
A;Reference number: JN0583; MUID:93203148; PMID:1284247
A;Accession: JN0583
A;Molecule type: mRNA
A;Residues: 1-1176 <KOB>
A;Cross-references: GB:S57131; NID:g298638; PIDN:AAB25794.1; PID:g298639
A;Experimental source: stomach
A;Accession: PN0490
A;Molecule type: protein
A;Residues: 44-55;721-728;828-851;1002-1019 <KO2>
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; fibronectin type II
C;Keywords: ATP; calmodulin binding; phosphotransferase
F;370-430/Domain: immunoglobulin homology <IMM1>
F;723-980/Domain: protein kinase homology <KIN>
F;731-739/Region: protein kinase ATP-binding motif
F;1083-1144/Domain: immunoglobulin homology <IMM2>

```
  Query Match             2.2%;  Score 160;  DB 2;  Length 1176;
  Best Local Similarity   18.1%;  Pred. No. 0.22;
  Matches  250;  Conservative  174;  Mismatches  470;  Indels  486;  Gaps   67;

Qy        134 PNNTP---GRTFGRTLSGPSGLSRPTST-----NTKRKDEP---ANVTFADPPK--RSLT 180
Db        121 PAETPKPLGNVKPAETPKPLGSTKPAETPKPLGSTKPAETPKPLGNVKPAETPKPLGNIK 180

Qy        181 RSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQCTKRAKGKLSDNVAAAAAP---- 236
Db        181 PTETPKPL-GSTKPAETPKPL--GSTKPAETPKPLGNVKPAETPKPLGNVKPAETPKPLG 237

Qy        237 -------PVPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACR 289
Db        238 NVKPAETPKPVSNA-------KPAETLKPVGNA-------------------------- 263

Qy        290 HNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDN-------QPLS 342
Db        264 -----------KPAETPKPLSNVKPA--ETPKLVGNAKPAETSKP-LDNAKPAEAPKPLG 309

Qy        343 SSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAA----RL 398
Db        310 NAKPAE------------------IPKPTGKEELKKEIKNDVNCKKGHAGATDSEKRPES 351

Qy        399 RNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLY 458
Db        352 RGTAPTFEEKLQDLHVA---EGQKLLLQCRV---------------SSDPPATITW---- 389

Qy        459 RLDAFRGKPFPEKPPNDVFVTAMTGNF---ESKGSAVVLSAVLDYNPDNSPTAPLY--LV 513
Db        390 ---TLNGK-----------TLKTTKFIVLSQEGSLCSVSIEKALPEDRG----LYKCVA 430

Qy        514 KLKPLMFEQGCRLTRRFGPDRFFEILIP-SPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQ 572
Db        431 KNSAGQAESSCQVT----------VDVPDAPTSENAKAPEMKARRPKSSLPPV----LGT 476

Qy        573 HSLVGRQWRAFFAKDAGY-RKPLREFQLRAEDPKPIIK----------ERVHFFAE---- 617
Db        477 ES------------DATVKKKPAPKTPPKAAMPPQIIQFPEDQKVRAGESVELFGKVAGT 524

Qy        618 -----TGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRI 672
Db        525 QPITCTWMKFRKQIQDSEHI--KVENSEQGSKLTI-----------RAARQEHCGCYTLL 571
```

```
Qy        673 ---QLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIR-DVLGL 728
Db        572 VENKLGSRQAQVNLTVV------DKPD--PPAGT-------------PCASDIRSSSLTL 610

Qy        729 GDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASE 788
Db        611 SWYGSSYDG--GSAVQSYSVEIWDSVDKTWKE------------------------LAT 643

Qy        789 LKSAGLNLQ-LLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYE 847
Db        644 CRSTSFNVQDLLP-----------------------DREYKFRVRAIN--------VYG 671

Qy        848 SYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLK 907
Db        672 --TSEPSQESE-----LTALGEKPEE----------EPKDEVEVSD---------DDEKE 705

Qy        908 SKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEES---------FTLLSDCDV 958
Db        706 PEVDYRTVTVNTEQKVSDFYDI-EERLGSGKFGQVFRLVEKKTGKIWAGKFFKAYS---- 760

Qy        959 LVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKL-SGGDYDGDMAWVCW 1017
Db        761 --AKEKENIRQEISIMNCLHHPKLVQCVDA--FEEKANIVMVLEIVSGG----------- 805

Qy       1018 DPEIVDGFVNAEMPL-EPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFA 1076
Db        806 --ELFERIIDEDFELTERECIKYMKQ----------------ISEGVEYIHKQGIVHLD 846

Qy       1077 LQPNFLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQ---LR 1133
Db        847 LKPENI-MCV----------NKTGTRIKLIDFGLARRLENAGSLKVLFGTPEFVAPEVIN 895

Qy       1134 RELLGGALSLPDPMYKSDSW-LGRGEPTHIIDYLKFSIARPAI-DKELEAFHNAMKAAKD 1191
Db        896 YEPIG---------YATDMWSIG------VICYILVSGLSPFMGDNDNETLANVTSATWD 940

Qy       1192 TEDGAHFWDPDLASYYTFFKEISDKSRS--SALLFTTLKNRIGEVE--------KEYGRL 1241
Db        941 FDDEA-------------FDEISDDAKDFISNLLKKDMKNRLNCTQCLQHPWLMKDTKNM 987

Qy       1242 VKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDKSG--ANYDSKVIRLLELSFLADREMN 1299
Db        988 EAKKLSKDRMKKYMAR-----RKW--------QKTGNAVRAIGRLSSMAMISGLSGRKSS 1034

Qy       1300 TWALLRASTAFKLYYHKSPKFVWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPD 1359
Db       1035 TGSPTSPLNAEKLESEDVSQAFLEAVAEEKPHVKP-----------------YFSKTIRD 1077

Qy       1360 KKFTKQYVARLEGDGSEYPDPEVY-----------------------------EVLGDDD 1390
Db       1078 LEVVEGSAARFDCKIEGYPDPEVVWFKDDQSIRESRHFQIDYDEDGNCSLIISDVCGDDD 1137
```

RESULT 15
H84582
hypothetical protein At2g19930 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: H84582
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: H84582
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-929 <STO>
A;Cross-references: GB:AE002093; NID:g3687225; PIDN:AAC62123.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g19930
A;Map position: 2

```
  Query Match             2.1%;  Score 159;  DB 2;  Length 929;
  Best Local Similarity   19.2%;  Pred. No. 0.17;
  Matches  179;  Conservative  140;  Mismatches  318;  Indels  293;  Gaps   46;

Qy        514 KLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQH 573
Db        193 RMGPLLEHTGTHLHKVLGDDNVLTVKFADVQKSSSTY---------SIDHYFTYKGIAKN 243

Qy        574 S-LVG-RQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVH-FFAETGITFRPDVFKTR 630
Db        244 GIMIGLRRYQFFVFKDGG----------KEEKKKDLSTKKVKCYFIRTDSTAFYDM-QNP 292

Qy        631 SVVPAEEPVEQRTEF----KVSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLE 686
Db        293 YILTGKSIYEARMHFMHVHRAPTLANYM---------------ARFSLILSKT---KTLE 334

Qy        687 P-------------HQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDV-- 731
Db        335 VDMTGITFDQIDDIHCHDQDGKDVLDKNKKPCIHSDGTGYISEDLARMCPLNIFKGKCLR 394

Qy        732 -----PSAVQGRF----GSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEV 782
Db        395 SESIQPLLIQFRMFYDGYAVKGTFLLN-----------------KKLC------PRTVQV 431

Qy        783 R-SVASELKSAGLN----LQLLPVLEDRARDKVKMRQAIGDRLIN--DLQRQFSEQKHAL 835
Db        432 RPSMIKVSKDPSLSNFSTFNALEVVTTRS-------QIFGAAALNYGEMDDQNAAQM--- 481

Qy        836 NRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIA 895
Db        482 ---------------------------ILVGIPLD--------------EPHLKNYLSIL- 500

Qy        896 WDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSD 955
Db        501 --LKTEKNDLKAGKL--PVTESYYLMGTVDPTGALKEDEVCVILESGQISGE-------- 548

Qy        956 CDVLVARSPA-HFPSDIQRVRAVFKPELH----SLKDVIIFSTKGDVPLAKKLSGGDYDG 1010
Db        549 --VLVYRNPGLHF-GDIHILKATYVKALEEYVGNSKFAVFFPQKGPRSLGDEIAGGDFDG 605

Qy       1011 DMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQ 1070
Db        606 DMYFISRNPELLENF----KPSEPWVS----------LTPPSKSNSGRAPSQLSPEELEE 651

Qy       1071 KSFHFALQPNF-----LGMCTN----YKERLCYINNSVSNKPAIILSSLVGNLVDQSKQG 1121
Db        652 ELFEMFLTAGFHASNVIGIAADSWLTIMDRFLILGDDRAEEKAEMKKKML-ELID----- 705

Qy       1122 IVFNEASWAQLRRELLGGALSLP--------------DPMYKSDSWLGRGEPTHIIDYLK 1167
Db        706 -IYYDALDAPKK----GDKVYLPNKLKPDIFPHYMERDKKFQSTSILGL-----IFDFVK 755

Qy       1168 FSIAR---PAIDKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLF 1224
Db        756 SQTTEEPSPSSGK---------------------GPLLVLLYT-YNQITKEIINHDMCF 792

Qy       1225 TTLKNRIGEVEK---------------------EYGRLVKNKEMRDSKDPYPVRVNQVYE 1263
Db        793 TVVYNL--EISKLPCFEDEPVSEFHMQKCRLWYDNYRTEMTQAMKTDKDESANEVIQRYK 850

Qy       1264 KWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYH---KSPKF 1320
Db        851 Q------EFYGAAGFE-DSK--KSLE-------ELYPQALALYKIVYDYAIHAGVSKCRF 894

Qy       1321 VWQMAGRQLA--YIKAQMTSRPGEGAPALM 1348
Db        895 VWKVAGPVLCRFYLNKKMQEKCLVCAPSVL 924
```

Search completed: March 29, 2004, 08:50:00
Job time : 35 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 29, 2004, 08:21:11 ; Search time 19 Seconds
(without alignments)
3842.229 Million cell updates/sec

Title: US-09-913-878A-2
Perfect score: 7397
Sequence: 1 MNPITPRKRNSPVEEIINRL..........YEVLGDDDFDGIGFTGNGDY 1402

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	182	2.5	1075	1	PST2_SCHPO	O13919 schizosacch
2	168	2.3	1226	1	SNT1_YEAST	P25357 saccharomyc
3	163.5	2.2	4377	1	ANK3_HUMAN	Q12955 homo sapien
4	160.5	2.2	1024	1	RIP3_MOUSE	P97434 mus musculu
5	160.5	2.2	5147	1	PCLO_HUMAN	Q9y6v0 homo sapien
6	160	2.2	1176	1	KMLS_BOVIN	Q28824 bos taurus
7	158.5	2.1	1736	1	ZO1_HUMAN	Q07157 homo sapien
8	157	2.1	1026	1	TAC2_HUMAN	O95359 homo sapien
9	156	2.1	2067	1	NCO6_MOUSE	Q9j119 m nuclear r
10	154.5	2.1	5120	1	PCLO_CHICK	Q9pu36 gallus gall
11	153	2.1	5430	1	MACF_HUMAN	Q9upn3 homo sapien
12	153	2.1	5938	1	MAC4_HUMAN	Q96pk2 homo sapien
13	151	2.0	1035	1	TAC2_MOUSE	Q9jjg0 mus musculu
14	151	2.0	3644	1	MINT_MOUSE	Q62504 mus musculu
15	148.5	2.0	1418	1	SMC4_YEAST	Q12267 saccharomyc
16	147.5	2.0	1385	1	FAT1_SCHPO	O13735 schizosacch
17	144	1.9	980	1	RIN3_MOUSE	P59729 mus musculu
18	144	1.9	2224	1	FA5_HUMAN	P12259 homo sapien
19	143.5	1.9	848	1	YBV8_YEAST	P38266 saccharomyc
20	143.5	1.9	860	1	CHI2_COCPO	P54197 coccidioide
21	141.5	1.9	3660	1	DMD_CHICK	P11533 gallus gall
22	140.5	1.9	1516	1	Y8I9_PSEAE	Q9hvt2 pseudomonas
23	140.5	1.9	3587	1	SRF2_BACSU	Q04747 bacillus su
24	140	1.9	1001	1	P5PA_RAT	Q9jmc1 rattus norv
25	139.5	1.9	1367	1	AMYH_YEAST	P08640 saccharomyc
26	139	1.9	3685	1	DMD_HUMAN	P11532 homo sapien
27	137	1.9	5171	1	BPEA_HUMAN	O94833 homo sapien
28	136.5	1.8	1403	1	YDF3_SCHPO	Q10475 schizosacch
29	136	1.8	1273	1	WEB1_YEAST	P38968 saccharomyc
30	135.5	1.8	956	1	YEF3_YEAST	P32618 saccharomyc
31	135.5	1.8	3969	1	HRX_HUMAN	Q03164 homo sapien
32	135	1.8	836	1	NOT3_YEAST	P06102 saccharomyc
33	135	1.8	877	1	WEE1_SCHPO	P07527 schizosacch
34	135	1.8	3396	1	PGCV_HUMAN	P13611 homo sapien
35	134.5	1.8	3588	1	SRF1_BACSU	P27206 bacillus su
36	133.5	1.8	1723	1	AIM1_HUMAN	Q9y4k1 homo sapien
37	133.5	1.8	1967	1	YG5O_YEAST	P53327 saccharomyc
38	133.5	1.8	2150	1	SDC3_CAEEL	P34706 caenorhabdi
39	133	1.8	1487	1	ICP4_HSVEK	P17473 equine herp
40	132.5	1.8	3343	1	BRC2_RAT	O35923 rattus norv
41	132	1.8	1395	1	IF4G_HUMAN	Q04637 homo sapien
42	132	1.8	1953	1	BNI1_YEAST	P41832 saccharomyc
43	132	1.8	2418	1	SPCA_HUMAN	P02549 homo sapien
44	132	1.8	3131	1	ESYN_FUSEQ	Q00869 fusarium eq
45	131.5	1.8	1337	1	DEXT_STRDO	P39653 streptococc

ALIGNMENTS

RESULT 1
PST2_SCHPO

```
ID   PST2_SCHPO     STANDARD;      PRT;  1075 AA.
AC   O13919;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Paired amphipathic helix protein pst2 (SIN3 homolog 2).
GN   PST2 OR SPAC23C11.15.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [2]
RP   GENE NAME.
RX   MEDLINE=99147069; PubMed=10022921;
RA   Dang V.D., Benedik M.J., Ekwall K., Choi J., Allshire R.C.,
RA   Levin H.L.;
RT   "A new member of the Sin3 family of corepressors is essential for
RT   cell viability and required for retroelement propagation in fission
RT   yeast.";
RL   Mol. Cell. Biol. 19:2351-2365(1999).
RN   [3]
RP   SEQUENCE OF 87-104; 280-291; 386-395; 424-446; 608-637; 678-709;
RP   780-788; 819-834; 869-883; 925-937 AND 1055-1075, FUNCTION, SUBUNIT,
```

```
RP   AND SUBCELLULAR LOCATION.
RX   MEDLINE=22658054; PubMed=12773392;
RA   Nakayama J.-I., Xiao G., Noma K.-I., Malikzay A., Bjerling P.,
RA   Ekwall K., Kobayashi R., Grewal S.I.S.;
RT   "Alp13, an MRG family protein, is a component of fission yeast Clr6
RT   histone deacetylase required for genomic integrity.";
RL   EMBO J. 22:2776-2787(2003).
CC   -!- FUNCTION: Has a role in chromatin assembly and chromosome
CC       segregation. Involved in the deacetylation of histones.
CC   -!- SUBUNIT: Heterotetramer of alp13, clr6, prw1 and pst2.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Contains 3 PAH (paired amphipathic helix) repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z98559; CAB11171.1; -.
DR   PIR; T38253; T38253.
DR   GeneDB_SPombe; SPAC23C11.15; -.
DR   InterPro; IPR003822; PAH.
DR   Pfam; PF02671; PAH; 3.
KW   Transcription regulation; Chromatin regulator; Nuclear protein;
KW   Repeat.
FT   REPEAT       54    100       PAH 1.
FT   REPEAT      160    206       PAH 2.
FT   REPEAT      271    317       PAH 3.
SQ   SEQUENCE   1075 AA;  124848 MW;  1AD301DB4EB1AFFB CRC64;

  Query Match             2.5%;  Score 182;  DB 1;  Length 1075;
  Best Local Similarity   18.7%;  Pred. No. 0.0061;
  Matches  193;  Conservative  157;  Mismatches  359;  Indels  322;  Gaps   51;

Qy        494 LSAVLDYNPDNSPTAPLYL-VKLKPLMFEQGCRLTRRFGPDRFFEILIPS-PTSTSPSVP 551
Db         81 ISVILRDYPDLLEYLNIFLPSSYKYLLSNSGANFTLQF--------TTPSGPVSTPSTYV 132

Qy        552 PVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQ------------ 598
Db        133 ATYNDLPCTYHRAIGFVSRVRRALLSNPEQFFKLQDS-----LRKFKNSECSLSELQTIV 187

Qy        599 --LRAEDPKPIIKERVHFFAETGITF--RPDV--FKTRSVVPAEEPVEQRTEFKVSQMLD 652
Db        188 TSLLAEHPS--LAHEFHNFLPSSIFFGSKPPLGSFPLRGI--------QSSQFTLSNISD 237

Qy        653 WLLQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVG 712
Db        238 LLSQ-----------------------------------SRPDNLSP--FSHLSNE--- 256

Qy        713 RMSRSVAKRIRDVL-----------------GLGDVPSAVQGRFG---SAKGMW--VID 749
Db        257 --SSDFFKNVKNVLTDVETYHEFLKLLNLYVQGIIDRNILVSRGFGFLKSNSGLWRSFLS 314

Qy        750 VDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDK 809
Db        315 LTSLSPEEFLSVY--------------------NSACSDFPECGPSYRLLPVEERNISCS 354

Qy        810 VKMRQAIGDRLIND----------------LQR---------QFSEQKHALNRPVEFRQW 844
Db        355 GRDDFAWG--ILNDDWVSHPTWASEESGFIVQRKTPYEEAMTKLEEERYEFDRHIEATSW 412

Qy        845 VYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKK-----------QKYLQD 893
Db        413 TIKSLKKIQNRINE--------LPEEERETYTLEEGLGLPSKSIYKKTIKLVYTSEHAEE 464

Qy        894 IAWDLQKRKCDTLK---SKLNIR------VGRSAYIYMIADFWGVLEENEVHVGFSS--- 941
Db        465 MFKALERMPCLTLPLVISRLEEKNEEWKSVKRS-----LQPGWRSIEFKNYDKSLDSQCV 519

Qy        942 --KFRDEE--ESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDV 997
Db        520 YFKARDKKNVSSKFLLAEADILRSQAKLHFP---LRSRSAFEFSFVYDNEIVLFDT---- 572

Qy        998 PLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTG 1057
Db        573 ------------------CY---MVCTYIVCNSP------SGLKKVEHFFKNILPLHFGL 605

Qy       1058 SAAKEQTTYDMIQKSFHFALQ-PNFLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVD 1116
Db        606 EKDKFSIFLDQVFRGPDYDVNAPNIVG------------NKPVRRKRSNSITQLT-EFVK 652

Qy       1117 QSK-QGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAI 1175
Db        653 QPKINGQRESRSAAAARKKEESGNKSQSNSQNSLSD------ESGNVTPVSKKQLSQPAA 706

Qy       1176 DKELEAFHNAMKA-----------------AKDTEDGAHFWDPD-----LASYYTFFKEI 1213
Db        707 ---------AIKASLKYPSHPDSLLEHQDHAGDTENEMHD-DVDKEQFGYSSMYVFFR-- 754

Qy       1214 SDKSRSSALLFTTLKNRIGEVEK--EYGRLVKNKEMRDSKDPYPVRVNQVY--EKWCAIT 1269
Db        755 ---------LFNLLYERLYELQRLEDQVSIIQQRII-----PNPVSQKQKIWRDRWNDLS 800

Qy       1270 PEAMDKSGANYDSKVIRLLELSF-LADREM--NTWALLRASTAFKLYYHKSPKFVWQMAG 1326
Db        801 DVPDEKT--HYENTYVMILRLIYGIVDQSAFEDYLRFYYGNKAYKIY--TIDKLVWS-AA 855

Qy       1327 RQLAYIKAQMTSRPGEGAPALMTAFM--YAGLMPDKKFTKQYVARLEGDGSEYPDPEVYE 1384
Db        856 KQVHHIVS-------DGKYKFVTSLVEQNSSASPKKNY-DDFLYRLEIEKLLNPDEILFR 907

Qy       1385 VLGDDDFDGIG 1395
Db        908 FCWINKFKSFG 918


RESULT 2
SNT1_YEAST
ID   SNT1_YEAST     STANDARD;      PRT;  1226 AA.
AC   P25357; Q02397; Q8NIL8;
DT   01-MAY-1992 (Rel. 22, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Potential DNA-binding protein SNT1.
GN   SNT1 OR YCR033W OR YCR33W OR YCR592.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92133166; PubMed=1776366;
RA   Wicksteed B.L., Roberts A.B., Sagliocco F.A., Brown A.J.P.;
RT   "The complete sequence of a 7.5 kb region of chromosome III from
RT   Saccharomyces cerevisiae that lies between CRY1 and MAT.";
RL   Yeast 7:761-772(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91335897; PubMed=1872032;
RA   Jia Y., Slonimski P.P., Herbert C.J.;
RT   "The complete sequence of the unit YCR59, situated between CRY1 and
RT   MAT, reveals two long open reading frames, which cover 91% of the
RT   10.1 kb segment.";
RL   Yeast 7:413-424(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=92244356; PubMed=1574125;
RA   Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
RA   Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
RA   Benit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
RA   Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
```

```
RA   Carignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,
RA   De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
RA   Dujon B., Duesterhoeft A., Erdmann D., Esteban M., Fabre F.,
RA   Fairhead C.A., Faye G., Feldmann H., Fiers W.,
RA   Francingues-Gaillard M.-C., Franco L., Frontali L., Fukuhara H.,
RA   Fuller L.J., Gent M.E., Gigot D., Gilliquet V., Glansdorff N.,
RA   Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haasemann M.,
RA   Hatat D., Hegemann J.H., Herbert C.J., Hilger F., Hohmann S.,
RA   Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
RA   Jacq C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
RA   Kleinhans U., Kreisl P., Lafranchi G., Lewis C., van der Linden C.G.,
RA   Lucchini G., Lutzenkirchen K., Maat C., Mannhaupt G., Manzano M.E.,
RA   Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
RA   Messenguy F., Mewes H.-W., Molemans F., Montague M.A., Navas L.,
RA   Newlon C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
RA   Perea J., Philippsen P., Pierard A., Planta R.J., Plevani P.,
RA   Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
RA   Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
RA   Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
RA   Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,
RA   Spiegelberg R., Stateva L.I., Steensma H.Y., Steiner S., Thierry A.,
RA   Thireos G., Triano L.N., Urrestarazu L.A., Valle G., Vetter I.,
RA   van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warmington J.R.,
RA   von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
RA   Zimmermann F.K., Sgouros J.G.;
RT   "The complete DNA sequence of yeast chromosome III.";
RL   Nature 357:38-46(1992).
RN   [4]
RP   REVISIONS.
RA   Valles G., Volckaerts G.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   IDENTIFICATION IN A COMPLEX WITH SET3; HST1; HOS2; SIF2; CPR1 AND
RP   HOS4.
RX   MEDLINE=21567937; PubMed=11711434;
RA   Pijnappel W.W.M.P., Schaft D., Roguev A., Shevchenko A., Tekotte H.,
RA   Wilm M., Rigaut G., Seraphin B., Aasland R., Stewart A.F.;
RT   "The S. cerevisiae SET3 complex includes two histone deacetylases,
RT   Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation
RT   gene program.";
RL   Genes Dev. 15:2991-3004(2001).
CC   -!- FUNCTION: Part of the Set3C complex, which is required to repress
CC       early/middle sporulation genes during meiosis.
CC   -!- SUBUNIT: Identified in a Set3C complex with SET3, HST1, HOS2,
CC       SIF2, CPR1 and HOS4.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: Contains 1 Myb-like domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; S78624; AAB21259.1; -.
DR   EMBL; X59075; CAA41799.1; -.
DR   EMBL; X59720; CAC42983.1; -.
DR   PIR; S15053; S15053.
DR   GermOnline; 138939; -.
DR   SGD; S0000629; SNT1.
DR   GO; GO:0000118; C:histone deacetylase complex; IDA.
DR   GO; GO:0003677; F:DNA binding; NAS.
DR   GO; GO:0017136; F:NAD-dependent histone deacetylase activity; IDA.
DR   GO; GO:0045129; F:NAD-independent histone deacetylase activity; IDA.
DR   GO; GO:0016575; P:histone deacetylation; IDA.
DR   GO; GO:0045835; P:negative regulation of meiosis; IDA.
DR   InterPro; IPR007137; DUF348.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   Pfam; PF03990; DUF348; 1.
DR   Pfam; PF00249; myb_DNA-binding; 2.
DR   SMART; SM00717; SANT; 2.
DR   PROSITE; PS50090; MYB_3; 1.
KW   Nuclear protein; DNA-binding; Coiled coil.
FT   DOMAIN       28    159       SER-RICH.
FT   DOMAIN      539    591       COILED COIL (POTENTIAL).
FT   DNA_BIND    884    934       MYB.
FT   DOMAIN      594    599       POLY-SER.
FT   CONFLICT    305    305       A -> T (IN REF. 2).
FT   CONFLICT    375    375       S -> N (IN REF. 2).
FT   CONFLICT    404    404       Q -> H (IN REF. 2).
FT   CONFLICT    435    435       V -> A (IN REF. 2).
FT   CONFLICT    442    442       N -> D (IN REF. 2).
FT   CONFLICT    482    482       E -> K (IN REF. 2).
FT   CONFLICT    513    513       L -> V (IN REF. 1).
FT   CONFLICT    644    644       H -> Y (IN REF. 2).
FT   CONFLICT    766    766       E -> D (IN REF. 2).
FT   CONFLICT    806    806       Q -> K (IN REF. 2).
FT   CONFLICT    831    831       V -> I (IN REF. 2).
FT   CONFLICT    846    846       G -> E (IN REF. 2).
FT   CONFLICT    851    851       D -> A (IN REF. 2).
FT   CONFLICT    855    857       AVQ -> GVR (IN REF. 2).
SQ   SEQUENCE   1226 AA;  138396 MW;  70AE6622227088EB CRC64;

  Query Match             2.3%;  Score 168;  DB 1;  Length 1226;
  Best Local Similarity   18.8%;  Pred. No. 0.051;
  Matches  221;  Conservative  174;  Mismatches  466;  Indels  314;  Gaps   50;

Qy         84 KAASSNWVPKAHADPDTLPWSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFG 143
Db         38 KSSNNGFVSSPTADNSTNP-SVTPSTAS------------------VPLPTAAPGSTFG 77

Qy        144 -----RTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFP 198
Db         78 IEAPRPSRYDPSSVSRPSSSSYSSTRKIGSRYNPDVERSSSTTSSTPESMNTSTITHTNT 137

Qy        199 DPVNTGSKRPSLESENLNQCTKRAKGKLSDNV-----AAAAAPPVPIASALDKVPTRRHA 253
Db        138 DIGNSRYSRKTMSRYN-------PQSTSSTNVTHFPSALSNAPPFYVANGSSRRP--RSM 188

Qy        254 NTRDPTATGHRRADQVDSFDTSQGTSYGS------SVFSACR-------HNQSTTQSSFE 300
Db        189 DDYSPDVTNKLETNNVSSVNNNSPHSYYSRSNKWRSIGTPSRPPFDNHVGNMTTTSNTNS 248

Qy        301 APPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQ--------PLSSSSQGETSFS 352
Db        249 IHQREPFWKAN-STTILKSTHSQSSPSL-HTKKFHDANKLDKPEASVKVETPSKDETKAI 306

Qy        353 TYYE-SFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWL-- 409
Db        307 SYHDNNFPPRKSVSKPNAPLEPDNIKVGEEDALGKKEVHKSG-----REIAKEHPTPVKM 361

Qy        410 --HEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKP 467
Db        362 KEHDELEARAKKVSKI-----------NIDGKQDEIWTTAKTVA---------SAVEVSK 401

Qy        468 FPEKPPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKL----------KP 517
Db        402 ESQK--------ELTRSVERKESPEIRDYERAYDPKALKTD---VTKLTVDNDNKSYEEP 450

Qy        518 LMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQ--HSL 575
Db        451 LEKVEGC--------------IFPLPKA-----------------ETRLWELKNQKRNKI 479

Qy        576 VGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPA 635
Db        480 ISEQKYLL-------KKAIRNF---SEYPFYAQNKLIHQQA-TGLI----LTKIISKIKK 524

Qy        636 EEPVEQRTEFKVSQMLDWL-LQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHK 694
Db        525 EEHLK-----KINLKHDYFDLQ---KKYEKECEILTKLSENLRKEEIENKRKEHELMEQK 576

Qy        695 TDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTG 754
Db        577 R----------REEGI---ETEKEKSLRH-------PSS------------------- 595
```

```
Qy      755 DEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRAR------- 807
                     : |:|:    ||||  :      :  :     |          : |:: |  |
Db      596 ------SSSSRRRNRADFVDDAEMENVLLQIDPNYKHYQAAATIPPLILDPIRKHSYKFC 649

Qy      808 ---DKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFL 864
                : |  ::     |:: |      |:: :|:|      | :  |   :   :  ::||    ::
Db      650 DVNNLVTDKKLWASRILKDASDNFTDHEHSL-----FLEG-YLIHPKKFGKISH----YM 699

Qy      865 AGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIA 924
              ||    :|  |::     :  ||     :  :   |  |:   :  :|  · |   ||
Db      700 GGLRSPEECVLHY-----YRTKKTVNYKQLLIDKNKKRKMSAAAKRRKRKERS------- 747

Qy      925 DFWGVLEENEVHVGFSSK-----FRDEEESFTLLSD--CDVLVARS---------PAHFP 968
                      : |  |  |  :          ||:|       :     |||   |              |
Db      748 ------NDEEVEVDESKEESTNTIEKEEKSENNAEENVQPVLVQGSEVKGDPLGTPEKVE 801

Qy      969 SDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPE--IVDGFV 1026
              : |::     |   ||  :  :   |            |        :          : :        ||:   :  :|:
Db      802 NMIEQRGEEFAGELENAERVNDLKRAHDEVGEESNKSSVIETNNGVQIMDPKGAVQNGYY 861

Qy     1027 NAE-----MPLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHF------ 1075
                   |          ||   |   |    |||                   ||  :  :|:|    :::|    |
Db      862 PEETKELDFSLENALQR--KKHK-------------SAPEHKTSYWSVRESQLFPELLKE 906

Qy     1076 -----ALQPNFLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWA 1130
                     :|     ||    :     |    |    |:    |     :::            |    :   |
Db      907 FGSQWSLISEKLGTKSTTMVRNYYQRNAARNGWKLLVDE-----TDLKRDGTSSESVQQS 961

Qy     1131 QL----RRELLGGALSLPDPMYKSDSWLGRGEPTH 1161
              |:       |    :      ::|  |    :           |:|||
Db      962 QILIQPERPNINAYSNIP-PQQRPALGYFVGQPTH 995
```

RESULT 3
ANK3_HUMAN

```
ID   ANK3_HUMAN     STANDARD;      PRT;  4377 AA.
AC   Q12955;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Ankyrin 3 (ANK-3) (Ankyrin G).
GN   ANK3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain stem;
RX   MEDLINE=95138209; PubMed=7836469;
RA   Kordeli E., Lambert S., Bennett V.;
RT   "AnkyrinG. A new ankyrin gene with neural-specific isoforms localized
RT   at the axonal initial segment and node of Ranvier.";
RL   J. Biol. Chem. 270:2352-2359(1995).
CC   -!- FUNCTION: Membrane-cytoskeleton linker. The neural-specific
CC       isoforms may participate in the maintenance/targeting of ion
CC       channels and cell adhesion molecules at the nodes of Ranvier and
CC       axonal initial segments.
CC   -!- SUBUNIT: Neural-specific isoforms may be a constituent of a
CC       neurofascin/NRCAM/ankyrin G complex.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=1;
CC         Comment=A number of isoforms are produced;
CC       Name=1; Synonyms=480-kDa isoform;
CC         IsoId=Q12955-1; Sequence=Displayed;
CC   -!- TISSUE SPECIFICITY: Expressed in brain and other tissues. Isoform
CC       1 is neural-specific.
CC   -!- SIMILARITY: Contains 23 ANK repeats.
CC   -!- SIMILARITY: Contains 1 death domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U13616; AAA64834.1; -.
DR   PIR; A55575; A55575.
DR   HSSP; P55273; 1BI8.
DR   Genew; HGNC:494; ANK3.
DR   MIM; 600465; -.
DR   GO; GO:0006605; P:protein targeting; NAS.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000488; Death.
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00023; ank; 24.
DR   Pfam; PF00531; death; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 21.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 21.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
KW   Cytoskeleton; Alternative splicing; Repeat; ANK repeat.
FT   REPEAT       73    102       ANK 1.
FT   REPEAT      106    135       ANK 2.
FT   REPEAT      139    168       ANK 3.
FT   REPEAT      172    201       ANK 4.
FT   REPEAT      203    230       ANK 5.
FT   REPEAT      234    263       ANK 6.
FT   REPEAT      267    296       ANK 7.
FT   REPEAT      300    329       ANK 8.
FT   REPEAT      333    362       ANK 9.
FT   REPEAT      366    395       ANK 10.
FT   REPEAT      399    428       ANK 11.
FT   REPEAT      432    461       ANK 12.
FT   REPEAT      465    494       ANK 13.
FT   REPEAT      498    527       ANK 14.
FT   REPEAT      531    560       ANK 15.
FT   REPEAT      564    593       ANK 16.
FT   REPEAT      597    626       ANK 17.
FT   REPEAT      630    659       ANK 18.
FT   REPEAT      663    692       ANK 19.
FT   REPEAT      696    725       ANK 20.
FT   REPEAT      729    758       ANK 21.
FT   REPEAT      762    791       ANK 22.
FT   REPEAT      795    825       ANK 23.
FT   DOMAIN     1519   1898       SER-RICH.
FT   DOMAIN     4090   4174       DEATH.
SQ   SEQUENCE   4377 AA;  480399 MW;  F42379E55768B684 CRC64;

  Query Match             2.2%;  Score 163.5;  DB 1;  Length 4377;
  Best Local Similarity   17.3%;  Pred. No. 0.66;
  Matches  275;  Conservative  205;  Mismatches  538;  Indels  575;  Gaps   72;

Qy       30 CVADTTLTPHRRKELAESDEDFGRHDKIYRALNFLYWRKDDSLNQAEANFFIEAKAASSN 89
            |  : ||  |:::  ::  |::   :  |:      :      ::      |   :          ||      ::
Db     1430 CNLNITLPAHKKETESDQDDEIEKTDRRQSFASLALRKRYSYLTEPG---MIERSTGATR 1486

Qy       90 WVPKAHA---------------DPDTLPW------------SKEPPRAATAGQQWALQTV 122
             :|   ::                        |  :|                  |    | |:       |: : |
Db     1487 SLPTTYSYKPFFSTRPYQSWTTAPITVPGPAKSGFTSLSSSSSNTPSASPLKSIWSVST- 1545

Qy      123 LLEVLNRFMPPP-NNTPGRTFGRTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLTR 181
                     |  |  :|  |  :     ::     | :: |    :  :      |         :    |
Db     1546 ---------PSPIKSTLGASTTSSVKSISDVASPI-RSLRTMSSPIKTVVSQSPYNIQVS 1595

Qy      182 SAT---------GPPIHGAAIPLKF---PDPVNTGSK----------RPSLESENLNQCT 219
            | |              |: |  |          |    |||                        |:      |:|   :
Db     1596 SGTLARAPAVTEATPLKGLASNSTFSSRTSPVTTAGSLLERSSITMTPPASPKSNINMYS 1655
```

```
Qy       220 KRAKGKLSDNVAAAAAPPV---------PIASALD-----KVPTRRHANTRDPTATGHRR 265
                   |  ::  :|||  :          |: | :|     |:       ::     ||
Db      1656 SSLPFK---SIITSAAPLISSPLKSVVSPVKSRVDVISSAKITMASSLSSPVKQMPGHAE 1712

Qy       266 ADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIES 325
             |:   :    :   |:: :  |:    :    :          |     | : |  |   :|
Db      1713 VALVNGSISPLKYASSSTLINGCKATATLQEKISSATNS-------VSSVVSAATDTVEK 1765

Qy       326 PSKGRTT------KSHIDNQPLSSSSQGETSFSTYYESFPSS------------------ 361
                  |            : |::    |   :  |    |  |    |  |  ||
Db      1766 VFSTTTAMPFSPLRSYVSAAPSAFQSLRTPSASALYTSLGSSISATTSSVTSSIITVPVY 1825

Qy       362 GGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEVT 421
                  :|||:     |    |       :|    : :|:                 |     |     :    |
Db      1826 SVVNVLPEPALKK-LPDSNSFTKSAAALLSPIKTLTTET---------HPQP---HFSRT 1872

Qy       422 RLFMHCKVDLEDESLGLKYDPSWSTA----RDVTDIWKTLYRLDAF------RGKPFPEK 471
                 :    :  |     :|  |     |  |::      :||  :: :  |  |:  |         |||   :
Db      1873 SSPVKSSLFLAPSALKLSTPSSLSSSQEILKDVAEMKEDLMRMTAILQTDVPEEKPFQPE 1932

Qy       472 PPNDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTR--- 528
             |  :          |   :  :      ::   |  :                |||:  :: :   |     :
Db      1933 LPKE-------GRIDDEEPFKIVEKVKE-----------DLVKVSEILKKDVCVDNKGSP 1974

Qy       529 ---------------RFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEE-------VI 565
                                |   :    |     :    |  |||:|  |    :    |      |       ||
Db      1975 KSPKSDKGHSPEDDWIEFSSEEIREARQQAAASQSPSLPERVQVKAKAASEKDYNLTKVI 2034

Qy       566 QWLT--MGQHSLVGRQWRAFFAKDAG------YRKP---LREFQLR--------AEDPKP 606
              :||   :|    ||    :::     ||    |          ||       |:|  :|:          :     |
Db      2035 DYLTNDIGSSSLTNLKYKFEDAKKDGEGGQKRVLKPAIALQEHKLKMPPASMRTSTSEKE 2094

Qy       607 IIKERVHFFAETGITFRPD--------------VFKTRSVVPAEEPVEQRTEFKVSQML 651
             :  |        ||     ||    ||                   |:|||     |      |
Db      2095 LCKMADSFFGTDTILESPDDFSQHDQDKSPLSDSGFETRSEKTPSAP--QSAE------- 2145

Qy       652 DWLLQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLL------SPSGTGE 705
                      |   |   ||   :  :                  |     |   :|:::         |
Db      2146 ---------TTGPK-PLFHEVPI------------PPVITETRTEVVHVIRSYDPSA--- 2180

Qy       706 VMNDGVGRMSRSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQ 765
                                          ||||                      |   |:  :    ||
Db      2181 ----------------------GDVP-------------------QTQPEEPVSPKPSP 2198

Qy       766 RKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQ 825
                  |            || :     |  :|                    :|||   |                 ::
Db      2199 TFME---------LEPKPTTSSIK----------------EKVKAFQ---------MK 2222

Qy       826 RQFSEQKH--ALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNF-----L 878
               |      |     |:: :  :  ::    |:: :   ||: :   |          :    |||::
Db      2223 ASSEEDDHNRVLSKGMRVKE---ETHITTTTRMVYHSPPGGEGASERIEETMSVHDIMKA 2279

Qy       879 MNSGFDPKKQ-----KYLQDIAWDLQKRKCDT----------LKSKLNIRVGRSAYIYMI 923
              || || |:      ::     :: |:  |    :|                :| ||     :
Db      2280 FQSGRDPSKELAGLFEHKSAVSPDVHKSAAETSAQHAEKDNQMKPKLERII--------- 2330

Qy       924 ADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELH 983
                          |||:              |              :|::   :    |                   ||
Db      2331 ---------EVHI---------EKGNQAEPTEVIIRETKKH------------PE-- 2355

Qy       984 SLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKD 1043
              |:: ::              | || ||                                  :|   :|  :
Db      2356 --KEMYVYQ--------KDLSRGDI-------------------------NLKDFLPEK 2379

Qy      1044 KTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNSVSNKP 1103
                 |        |    |     :|:  |  :                                     :  |
Db      2380 HDAFP---CSEEQGQQEEEELTAE--------------------------------ESLP 2404

Qy      1104 AIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHII 1163
              : : ||  |     |  |  :       :   |  |||             :  |    ||:     |  :      |  |
Db      2405 SYLESSRVNTPVSQEED----SRPSSAQL----------ISDDSYKTLKLLSQ----HSI 2446

Qy      1164 DY-----------LKFSIARPAIDKELEAFHNAMKAAKDTEDGA--HFWDPDLASYYTF 1209
                     :|                  :|:  :   :  ::|:  |:      |||:      |    |
Db      2447 EYHDDELSELRGESYRFA-EKMLLSEKLDVSHS------DTEESVTDHAGPPS------- 2492

Qy      1210 FKEI--SDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQV------ 1261
                  |:   ||| ||     :   |    |     |    :  |   |   : :|      :   |||   :    :|    :
Db      2493 -SELQGSDK-RSREKIATAPKKEI--LSKIYKDVSENGVGKVSKDEHFDKVTVLHYSGNV 2548

Qy      1262 ----YEKWCAITPEAMDKSGAN--YDSKVIR--------LLELS-FLADR------EMNT 1300
                 :    |    |  :  :|            |: :|  |             | |:|  |  |  |:         |: :
Db      2549 SSPKHAMWMRFTEDRLDRGREKLIYEDRVDRTVKEAEEKLTEVSQFFRDKTEKLNDELQS 2608

Qy      1301 WALLRASTAFKLYYHKSPK-------FVWQMAGRQLAYIKAQMTSRPGEG-------APA 1346
                                  |  |   :||                :  ::          ::||:          |:|           ||:
Db      2609 PEKKARPKNGKEYSSQSPTSSSPEKVLLTELLASNDEWVKARQHGPDGQGFPKAEEKAPS 2668

Qy      1347 LMTAFMYAGLMPDKKFTKQYV----ARLEGDGS 1375
              | ::         |:|      |          : :|  ||
Db      2669 LPSS-------PEKMVLSQQTEDSKSTVEAKGS 2694
```

RESULT 4
RIP3_MOUSE

```
ID   RIP3_MOUSE     STANDARD;      PRT;  1024 AA.
AC   P97434;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Rho-interacting protein 3 (p116RIP) (RIP3).
GN   RHOIP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97344280; PubMed=9199174;
RA   Gebbink M.F.B.G., Kranenburg O., Poland M., van Horck F.P.G.,
RA   Houssa B., Moolenaar W.H.;
RT   "Identification of a novel, putative Rho-specific GDP/GTP exchange
RT   factor and a RhoA-binding protein: control of neuronal morphology.";
RL   J. Cell Biol. 137:1603-1613(1997).
CC   -!- FUNCTION: Rhoa-binding protein involved in control of the actin
CC       cytoskeleton. Overexpression promotes neuronal cell flattening and
CC       neurite outgrowth, probably by counteracting rhoa-mediated
CC       signaling.
CC   -!- TISSUE SPECIFICITY: Highly enriched in the brain.
CC   -!- SIMILARITY: Contains 2 PH domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U73200; AAB18198.1; -.
DR   PIR; T30868; T30868.
DR   MGD; MGI:1349438; Rhoip3.
DR   InterPro; IPR001849; PH.
DR   Pfam; PF00169; PH; 2.
DR   SMART; SM00233; PH; 2.
DR   PROSITE; PS50003; PH_DOMAIN; 2.
KW   Guanine-nucleotide releasing factor; Repeat; Coiled coil.
FT   DOMAIN       43    150       PH 1.
FT   DOMAIN      386    482       PH 2.
FT   DOMAIN      164    172       PRO-RICH.
FT   DOMAIN      185    190       POLY-SER.
FT   DOMAIN      285    296       PRO-RICH.
FT   DOMAIN      671    976       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1024 AA;  116363 MW;  A91BDBB955E45E02 CRC64;
```

```
  Query Match             2.2%;  Score 160.5;  DB 1;  Length 1024;
  Best Local Similarity   19.9%;  Pred. No. 0.11;
  Matches  211;  Conservative  153;  Mismatches  407;  Indels  291;  Gaps   51;

Qy        161 KRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQCTK 220
Db        159 KRKVEPPTPQEPGPAKMAVTSSSGGTSGSSSSIP--------SAEKVPTTKS-TLWQEEM 209

Qy        221 RAKGKLSDNVAAAAAPP---VPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQG 277
Db        210 RAKDQPDGTSLSPAQSPSQSQPPAACTPREP-GLESKEDESTISG-------DRVDGGRK 261

Qy        278 TSYGSSVFSACRHNQSTTQSSFEAP------PSQPREKRPVDATVFEAGHLIESPSKGRT 331
Db        262 VRVESGYFSLEKAKQDLRAEEQLPPLLSPPSPSTPHSRRSQVIEKFEA--LDIEKAEHME 319

Qy        332 TKSHIDNQPLSSSSQGET---------SFSTYYESFPSSGGEGAIPEPSRSNGLARSEES 382
Db        320 TNMLILTTPSSDTRQGRSERRAIPRKRDFASEAPTAPLSDACPLSPHRRAKSLDRRSTES 379

Qy        383 ARSQVQVHAPVVAARLRNIWPKFPK-WLHEAPLAVAWE------VTRLFMHCKVDLEDES 435
Db        380 S---------MTPDLLN----FKKGWLTKQYEDGQWKKHWFVLADQSLRYYRDSVAEEA 425

Qy        436 LGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLS 495
Db        426 ADLDGEINLSTCYDVTE-----YPVQRNYGFQIHTK-EGEFTLSAMTSGIRRNWIQTIMK 479

Qy        496 AVLDYNPDNSP--TAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPV 553
Db        480 HVL---PASAPDVTSSLPEGKNKSTSFETCSRSTEK--------------QEAEPGEPDP 522

Qy        554 VSKQPGAVE-------EVIQWLTMG--QHSLVGRQWRAFFAKDAGYRKPL--------RE 596
Db        523 EQKKSRARERRREGRSKTFDWAEFRPIQQALAQERASAVGSSDSGDPGCLEAEPGELERE 582

Qy        597 FQLRAEDPK-----------PIIKE---RVHFFAETGITFRPDVFKTRSV-VPAEE---- 637
Db        583 RARRREEPRKRFGMLDTIDGPGMEDTALRMDIDRSPGLLGTPDL-KTQNVHVEIEQRWHQ 641

Qy        638 ----PVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQ-LGLSKTYAIMTLEPHQIRH 692
Db        642 VETTPLREEKQVPIAPL--------------HLSLEDRSERLSTHELTSLLEKELEQSQK 687

Qy        693 HKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVD- 751
Db        688 EASDLL--------------EQNRLLQDQLRVAL----------GREQSAREGYVLQATC 723

Qy        752 DTGDEDWIETYPSQRKWECDFVDKHQRTLE-VRSVASEL--KSAGLNLQLLPVLEDRARD 808
Db        724 ERGFAAMEETH--QKKIE-DLQRQHQRELEKLREEKDRLLAEETAATISAIEAMKNAHRE 780

Qy        809 KVKMRQAIGDRL--------INDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGR 860
Db        781 EMERELEKSQRSQISSINSDIEALRRQYLEELQSVQRELEV---LSEQYSQKCLENAH-- 835

Qy        861 VPFLAGLPDSQEETLNFLM--NSGFDPKKQKYLQDIAWDLQKRKCDTL--------KSKL 910
Db        836 ---LAQALEAERQALRQCQRENQELNAHNQELNNRLAAEITRLR--TLLTGDGGGESTGL 890

Qy        911 NIRVGRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSD 970
Db        891 PLTQGKDAY---------------------------------ELEVLLRVK----ESE 911

Qy        971 IQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEM 1030
Db        912 IQ----YLKQEISSLKDELQTALRD-----KKYASDKY----------KDIYTELSIAKA 952

Qy       1031 PLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTT---YDMIQKSFHFALQPNFLGMCTN 1087
Db        953 KADCDISRLKEQLKA------ATEALGEKSPEGTTVSGYDIMKSKSN----PDFL----- 997

Qy       1088 YKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASW 1129

Db        998 KKDRSC-------------VTRQLRNIRSKS----VIEQVSW 1022

RESULT 5
PCLO_HUMAN
ID   PCLO_HUMAN     STANDARD;      PRT;  5147 AA.
AC   Q9Y6V0; O43373; O60305; Q9BVC8; Q9UIV2; Q9Y6U9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Piccolo protein (Aczonin) (Fragments).
GN   PCLO OR ACZ OR KIAA0559.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-759 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with rim and bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
RN   [2]
RP   SEQUENCE OF 552-4404 FROM N.A.
RA   Kraemer J., Wollam C., Wohldmann P., McGrane B.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=98290545; PubMed=9628581;
RA   Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:31-39(1998).
RN   [4]
RP   SEQUENCE OF 4405-4439 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SEQUENCE OF 4405-5147 FROM N.A.
RA   Kalicki J., Elliott G.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
CC       trafficking (By similarity).
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1, RIMS2 and profilin (By
```

```
CC        similarity).
CC    -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC        synaptic junctions (By similarity).
CC    -!- ALTERNATIVE PRODUCTS:
CC        Event=Alternative splicing; Named isoforms=2;
CC          Comment=Additional isoforms seem to exist;
CC        Name=1;
CC          IsoId=Q9Y6V0-1; Sequence=Displayed;
CC        Name=2;
CC          IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
CC                                   VSP_003926, VSP_003927;
CC          Note=No experimental confirmation available;
CC    -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC        phospholipids. Calcium binds with low affinity but with high
CC        specificity and induces a large conformational change.
CC    -!- SIMILARITY: Contains 2 C2 domains.
CC    -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC    the European Bioinformatics Institute.  There are no  restrictions on  its
CC    use  by  non-profit  institutions as long  as its content  is  in  no  way
CC    modified and this statement is not removed.  Usage  by  and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------------
DR    EMBL; Y19188; CAB60727.1; -.
DR    EMBL; AC004903; AAD20936.1; -.
DR    EMBL; AC004886; AAD21789.1; -.
DR    EMBL; AB011131; BAA25485.1; -.
DR    EMBL; BC001304; AAH01304.1; -.
DR    EMBL; AC004082; AAB97937.1; -.
DR    PIR; T00634; T00634.
DR    HSSP; P04410; 1A25.
DR    Genew; HGNC:13406; PCLO.
DR    MIM; 604918; -.
DR    GO; GO:0005856; C:cytoskeleton; NAS.
DR    GO; GO:0045202; C:synaptic junction; ISS.
DR    GO; GO:0005509; F:calcium ion binding; ISS.
DR    GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
DR    GO; GO:0005522; F:profilin binding; ISS.
DR    GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR    GO; GO:0016080; P:synaptic vesicle targeting; ISS.
DR    InterPro; IPR000008; C2.
DR    InterPro; IPR001565; Synaptotagmin.
DR    PRINTS; PR00360; C2DOMAIN.
DR    PRINTS; PR00399; SYNAPTOTAGMN.
DR    SMART; SM00239; C2; 2.
DR    PROSITE; PS00499; C2_DOMAIN_1; 1.
DR    PROSITE; PS50004; C2_DOMAIN_2; 2.
KW    Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
KW    Repeat; Alternative splicing.
FT    NON_TER       1      1
FT    DOMAIN      400    465       10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                 P-A-K-P-Q-P-Q-Q-P-X.
FT    ZN_FING     499    523       C4-TYPE (POTENTIAL).
FT    ZN_FING     969    992       C4-TYPE (POTENTIAL).
FT    NON_CONS   1010   1011
FT    DOMAIN     2300   2325       POLY-PRO.
FT    DOMAIN     4391   4442       PDZ.
FT    DOMAIN     4544   4633       C2 DOMAIN 1.
FT    DOMAIN     5031   5121       C2 DOMAIN 2.
FT    VARSPLIC   4404   4404       S -> SGNGLGIRIVGGKEIPGHSGEIGAYIAKILPGGSAE
FT                                 QTGKLMEG (in isoform 2).
FT                                 /FTId=VSP_003923.
FT    VARSPLIC   4534   4534       K -> KPTDGTKVVSHPITGEIQ (in isoform 2).
FT                                 /FTId=VSP_003924.
FT    VARSPLIC   4576   4576       G -> GQVMVVQNAS (in isoform 2).
FT                                 /FTId=VSP_003925.
FT    VARSPLIC   4757   4761       TAHKS -> SKRRK (in isoform 2).
FT                                 /FTId=VSP_003926.
FT    VARSPLIC   4762   5147       Missing (in isoform 2).
FT                                 /FTId=VSP_003927.
```

```
SQ   SEQUENCE   5147 AA;  563537 MW;  CD5D84990498CD3C CRC64;

  Query Match             2.2%;  Score 160.5;  DB 1;  Length 5147;
  Best Local Similarity   18.8%;  Pred. No. 1.3;
  Matches  286;  Conservative  195;  Mismatches  539;  Indels  505;  Gaps   72;

Qy         92 PKAHADPDTLPWSKEPPRAATAGQQWALQTVLLEVLNRFMPPPNNTPG----------- 139
Db        620 PTLHGSPSAK--AKQPPEADSLSK----------------PAPPKEPSVPSEQDKAPVAD 661

Qy        140 ---------RTFGRTLSGPSGLSRPTSTNTKRKDEPANVTF----ADPPKRSLTRSATG- 185
Db        662 DKPKQPKMVKPTTDLVSSSSATTKPDIPSSKVQSQAEEKTTPPLKTDSAKPSQSFPPTGE 721

Qy        186 --PPIHGAAIPLKFPDPVNTGSKRPSLESENLNQC--------TKRAKGKLSDNVAA--- 232
Db        722 KVTPFDSKAIPRPASDSKIISHPGPSSESKGQKQVDPVQKKEEPKKAQTKMSPKPDAKPM 781

Qy        233 ---AAAPPVPIASALDKVPTRRHANTRDPTATGHRRADQVDSF--------DTSQGT--- 278
Db        782 PKGSPTPPGPRPTAGQTVPTPQ--QSPKPQEQSRRFSLNLGSITDAPKSQPTTPQETVTG 839

Qy        279 ---SYGSSVFSACRHNQST-------TQSSFEAP------PSQ-PREKRPVDATVFEAGH 321
Db        840 KLFGFGASIFSQASNLISTAGQPGPHSQSGPGAPMKQAPAPSQPPTSQGPPKST----GQ 895

Qy        322 LIESPSKG----RTTKSHIDNQPLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLA 377
Db        896 APPAPAKSIPVKKETKAPAAEKLEPKAEQAPTVKRTETEKKP--------PPIKDSKSLT 947

Qy        378 RSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLG 437
Db        948 AEPQKA---------VLPTKLE----KSPKPESTCPL------------CKTEL---NIG 979

Qy        438 LKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSAV 497
Db        980 SKDPPNFNTC---TECKNQVCNLCGF--NPTPHLTENCQTQRAISGQ---------LGDI 1025

Qy        498 LDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTS---PSVPPVV 554
Db       1026 RKMPP--APSGP----KASPM---------------------PVPTESSSQKTAVPPQV 1057

Qy        555 S--------------------KQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAG---- 589
Db       1058 KLVKKQEQEVKTEAEKVILEKVKETLSMEKIPPMVTTDQ-----KQEESKLEKDKASALQ 1112

Qy        590 YRKPL--------REFQLRAEDPKPIIKERVHFFAETGI-----TFRP-----DVFKTRS 631
Db       1113 EKKPLPEEKKLIPEEEKIRSEEKKPLLEEKKPTPEDKKLLPEAKTSAPEEQKHDLLKSQ- 1171

Qy        632 VVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLK---LFSRIQLGLSK-----TYAI- 682
Db       1172 VQIAEEKLEGRVAPKTVQ----------EGKQPQTKMEGLPSGTPQSLPKEDDKTTKTIK 1221

Qy        683 ------MTLEPHQIRH-HKTDL------LSPSGTGEV--MNDGVGRMSRSVAKRIRDVLG 727
Db       1222 EQPQPPCTAKPDQEKEDDKSDTSSSQQPKSPQGLSDTGYSSDGISS------------S 1268

Qy        728 LGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVAS 787
Db       1269 LGEIPSLI----------------PTDEKDILKGL------------KKDSFSQESSPSS 1300

Qy        788 ELKSAGLNLQLLPVLEDRARDKV--KMRQAIGDRLINDLQRQFSEQKHALNRPVEF---R 842
Db       1301 PSDLAKLESTVLSILEAQASTLADEKSEKKTQPHEVSPEQPKDQEKTQSLSETLEITISE 1360

Qy        843 QWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIA-----WD 897
Db       1361 EEIKESQEERKDTFK----------KDSQQD----IPSSKDHKEKSEFVDDITTRREPYD 1406

Qy        898 LQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSS----KFRD-----EEE 948
Db       1407 SVEESSESENSPVPQRKRRTS------------------VGSSSSDEYKQEDSQGSGEEE 1448
```

```
Qy        949 SFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDY 1008
                |  :      :  |    |     |  : :|    |     |  :        |||      |:: |  :
Db       1449 DF-IRKQIIEMSADEDASGSEDDEFIRNQLKEISSSTESQKKEETKG----KGKITAGKH 1503

Qy       1009 -----------DGDM----AWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMAS 1053
                         | |      :|   | |  |              |:| :| :       ::|: :
Db       1504 RRLTRKSSTSIDEDAGRRHSWHDEDDEAFDE--------SPEL-KYRETKSQESEELVVT 1554

Qy       1054 HGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNSVSNKPAIILSSLVGN 1113
               | |                                      : :   :|:::::| :   |
Db       1555 GGGGLR----------------------------RFKTIELNSTIADKYSAESS----- 1580

Qy       1114 LVDQSKQGIVFNEASWAQLRRELLGGALSLPD-PMYKSDSWLGRGEPTHIIDYLKFS--- 1169
                  | |  : |:|    :|   |      || |  |  :|     ||| :   :     :
Db       1581 ---QKKTSLYFDEE--PELEME------SLTDSPEDRS-----RGEGSSSLHASSFTPGT 1624

Qy       1170 --IARPAIDKELEAFHNAMKA-AKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTT 1226
                :  ::|:: ::  :   |  :|        |     |          |  :      ||
Db       1625 SPTSVSSLDEDSDSSPSHKKGESKQQRKARHRPHGPLLPTIEDSSEEEELREEEELLKEQ 1684

Qy       1227 LKNRIGEVEKEYGRLVKNKEMRDS----------------------KDPYP-------VR 1257
               | |  |:|::  :       |  :|                     :|  |
Db       1685 EKQR--EIEQQQRKSSSKKSKKDKDELRAQRRRERPKTPPSNLSPIEDASPTEELRQAAE 1742

Qy       1258 VNQVYEKWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHKS 1317
                : :::     |:      :::      ::     :::|:               : :  || ||
Db       1743 MEELHRSSCSEYSPSIESDPEGFEISPEKIIEVQ----------KVYKLPTAVSLYSPTD 1792

Qy       1318 PKFVWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGD---- 1373
                : :  |  | |  |    |:                   ||  :|       | :  |  ||
Db       1793 EQSIMQKEGSQKALKSAEE---------------MYEEMMHKTHKYKAFPAANERDEVFE 1837

Qy       1374 -----GSEYPDPEVYEVLGDDDFDG 1393
                      |    :   :||  | :| ::|
Db       1838 KEPLYGGMLIEDYIYESLVEDTYNG 1862
```

```
RESULT 6
KMLS_BOVIN
ID   KMLS_BOVIN     STANDARD;      PRT;  1176 AA.
AC   Q28824;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Myosin light chain kinase, smooth muscle (EC 2.7.1.117) (MLCK)
DE   [Contains: Telokin].
GN   MYLK.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Stomach;
RX   MEDLINE=93203148; PubMed=1284247;
RA   Kobayashi H., Inoue A., Mikawa T., Kuwayama H., Hotta Y., Masaki T.,
RA   Ebashi S.;
RT   "Isolation of cDNA for bovine stomach 155 kDa protein exhibiting
RT   myosin light chain kinase activity.";
RL   J. Biochem. 112:786-791(1992).
CC   -!- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR
CC       SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE
CC       IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT
CC       FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS.
CC   -!- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
CC       light-chain] phosphate.
CC   -!- SUBUNIT: ISOFORM TELOKIN BINDS CALMODULIN.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative initiation;
CC         Comment=At least 3 isoforms, Non-muscle, Smooth-muscle (shown
CC         here) and Telokin, are produced by alternative initiation.
CC         Isoform Non-muscle is the longest and telokin is a C-terminal
CC         section without catalytic activity;
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC   -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; S57131; AAB25794.1; -.
DR   PIR; JN0583; JN0583.
DR   HSSP; P56276; 1TLK.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 2.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS50835; IG_LIKE; 3.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Transferase; Serine/threonine-protein kinase; Calmodulin-binding;
KW   ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
KW   Alternative initiation.
FT   CHAIN           1   1176       MYOSIN LIGHT CHAIN KINASE, ISOFORM
FT                                  SMOOTH-MUSCLE.
FT   CHAIN        1022   1176       MYOSIN LIGHT CHAIN KINASE, ISOFORM
FT                                  TELOKIN.
FT   INIT_MET     1022   1022       FOR ISOFORM TELOKIN.
FT   DOMAIN        100    291       16 X 12 AA TANDEM REPEATS.
FT   REPEAT        100    111       1.
FT   REPEAT        112    123       2.
FT   REPEAT        124    135       3 (INCOMPLETE).
FT   REPEAT        136    147       4.
FT   REPEAT        148    159       5.
FT   REPEAT        160    171       6.
FT   REPEAT        172    183       7.
FT   REPEAT        184    195       8.
FT   REPEAT        196    207       9.
FT   REPEAT        208    219       10.
FT   REPEAT        220    231       11.
FT   REPEAT        232    243       12.
FT   REPEAT        244    255       13.
FT   REPEAT        256    267       14.
FT   REPEAT        268    279       15.
FT   REPEAT        280    291       16.
FT   DOMAIN        356    444       IG-LIKE C2-TYPE 1.
FT   DOMAIN        498    586       IG-LIKE C2-TYPE 2.
FT   DOMAIN        603    673       FIBRONECTIN TYPE-III.
FT   DOMAIN        725    980       PROTEIN KINASE.
FT   DOMAIN        972   1035       CALMODULIN-BINDING.
FT   DOMAIN       1069   1158       IG-LIKE C2-TYPE 3.
FT   NP_BIND       731    739       ATP (BY SIMILARITY).
FT   BINDING       754    754       ATP (BY SIMILARITY).
FT   ACT_SITE      846    846       BY SIMILARITY.
FT   DOMAIN       1171   1176       POLY-GLU.
SQ   SEQUENCE   1176 AA;  128824 MW;  F53DC6D4D42D4B97 CRC64;
```

```
Query Match            2.2%;  Score 160;  DB 1;  Length 1176;
Best Local Similarity  18.1%;  Pred. No. 0.14;
Matches  250;  Conservative  174;  Mismatches  470;  Indels  486;  Gaps   67;

Qy    134 PNNTP---GRTFGRTLSGPSGLSRPTST-----NTKRKDEP---ANVTFADPPK--RSLT 180
Db    121 PAETPKPLGNVKPAETPKPLGSTKPAETPKPLGSTKPAETPKPLGNVKPAETPKPLGNIK 180

Qy    181 RSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQCTKRAKGKLSDNVAAAAAP---- 236
Db    181 PTETPKPL-GSTKPAETPKPL--GSTKPAETPKPLGNVKPAETPKPLGNVKPAETPKPLG 237

Qy    237 -------PVPIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACR 289
Db    238 NVKPAETPKPVSNA-------KPAETLKPVGNA-------------------------- 263

Qy    290 HNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDN-------QPLS 342
Db    264 -----------KPAETPKPLSNVKPA--ETPKLVGNAKPAETSKP-LDNAKPAEAPKPLG 309

Qy    343 SSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAA----RL 398
Db    310 NAKPAE------------------IPKPTGKEELKKEIKNDVNCKKGHAGATDSEKRPES 351

Qy    399 RNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLY 458
Db    352 RGTAPTFEEKLQDLHVA---EGQKLLLQCRV---------------SSDPPATITW---- 389

Qy    459 RLDAFRGKPFPEKPPNDVFVTAMTGNF---ESKGSAVVLSAVLDYNPDNSPTAPLY--LV 513
Db    390 ---TLNGK-----------TLKTTKFIVLSQEGSLCSVSIEKALPEDRG----LYKCVA 430

Qy    514 KLKPLMFEQGCRLTRRFGPDRFFEILIP-SPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQ 572
Db    431 KNSAGQAESSCQVT---------VDVPDAPTSENAKAPEMKARRPKSSLPPV----LGT 476

Qy    573 HSLVGRQWRAFFAKDAGY-RKPLREFQLRAEDPKPIIK-----------ERVHFFAE---- 617
Db    477 ES------------DATVKKKPAPKTPPKAAMPPQIIQFPEDQKVRAGESVELFGKVAGT 524

Qy    618 -----TGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRI 672
Db    525 QPITCTWMKFRKQIQDSEHI--KVENSEQGSKLTI-----------RAARQEHCGCYTLL 571

Qy    673 ---QLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIR-DVLGL 728
Db    572 VENKLGSRQAQVNLTVV------DKPD--PPAGT-------------PCASDIRSSSLTL 610

Qy    729 GDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASE 788
Db    611 SWYGSSYDG--GSAVQSYSVEIWDSVDKTWKE-------------------------LAT 643

Qy    789 LKSAGLNLQ-LLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYE 847
Db    644 CRSTSFNVQDLLP-----------------------DREYKFRVRAIN--------VYG 671

Qy    848 SYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLK 907
Db    672 --TSEPSQESE-----LTALGEKPEE----------EPKDEVEVSD---------DDEKE 705

Qy    908 SKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEES---------FTLLSDCDV 958
Db    706 PEVDYRTVTVNTEQKVSDFYDI-EERLGSGKFGQVFRLVEKKTGKIWAGKFFKAYS---- 760

Qy    959 LVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKL-SGGDYDGDMAWVCW 1017
Db    761 --AKEKENIRQEISIMNCLHHPKLVQCVDA--FEEKANIVMVLEIVSGG----------- 805

Qy   1018 DPEIVDGFVNAEMPL-EPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFA 1076
Db    806 --ELFERIIDEDFELTERECIKYMKQ-----------------ISEGVEYIHKQGIVHLD 846

Qy   1077 LQPNFLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQ---LR 1133
Db    847 LKPENI-MCV----------NKTGTRIKLIDFGLARRLENAGSLKVLFGTPEFVAPEVIN 895

Qy   1134 RELLGGALSLPDPMYKSDSW-LGRGEPTHIIDYLKFSIARPAI-DKELEAFHNAMKAAKD 1191
Db    896 YEPIG---------YATDMWSIG------VICYILVSGLSPFMGDNDNETLANVTSATWD 940

Qy   1192 TEDGAHFWDPDLASYYTFFKEISDKSRS--SALLFTTLKNRIGEVE--------KEYGRL 1241
Db    941 FDDEA-----------;-FDEISDDAKDFISNLLKKDMKNRLNCTQCLQHPWLMKDTKNM 987

Qy   1242 VKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDKSG--ANYDSKVIRLLELSFLADREMN 1299
Db    988 EAKKLSKDRMKKYMAR-----RKW--------QKTGNAVRAIGRLSSMAMISGLSGRKSS 1034

Qy   1300 TWALLRASTAFKLYYHKSPKFVWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPD 1359
Db   1035 TGSPTSPLNAEKLESEDVSQAFLEAVAEEKPHVKP-----------------YFSKTIRD 1077

Qy   1360 KKFTKQYVARLEGDGSEYPDPEVY-----------------------------EVLGDDD 1390
Db   1078 LEVVEGSAARFDCKIEGYPDPEVVWFKDDQSIRESRHFQIDYDEDGNCSLIISDVCGDDD 1137

RESULT 7
ZO1_HUMAN
ID   ZO1_HUMAN      STANDARD;      PRT;  1736 AA.
AC   Q07157;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tight junction protein ZO-1 (Zonula occludens 1 protein) (Zona
DE   occludens 1 protein) (Tight junction protein 1).
GN   TJP1 OR ZO1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93361541; PubMed=8395056;
RA   Willott E., Balda M.S., Fanning A.S., Jameson B., van Itallie C.,
RA   Anderson J.M.;
RT   "The tight junction protein ZO-1 is homologous to the Drosophila
RT   discs-large tumor suppressor protein of septate junctions.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:7834-7838(1993).
RN   [2]
RP   INTERACTION WITH CGN.
RX   MEDLINE=22140336; PubMed=12023291;
RA   D'Atri F., Nadalutti F., Citi S.;
RT   "Evidence for a functional interaction between cingulin and ZO-1 in
RT   cultured cells.";
RL   J. Biol. Chem. 277:27757-27764(2002).
CC   -!- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL
CC       REQUIRED FOR TIGHT JUNCTION ASSEMBLY, WHILE THE C-TERMINAL MAY
CC       HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN
CC       MIGHT BE INVOLVED IN STABILIZING JUNCTIONS.
CC   -!- SUBUNIT: Interacts with occludin, claudins, TJP3/ZO-3 and
CC       cingulin.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.
CC       MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT
CC       OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=Q07157-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=Q07157-2; Sequence=VSP_003148;
CC   -!- TISSUE SPECIFICITY: THE ALPHA-CONTAINING ISOFORM IS FOUND IN MOST
CC       EPITHELIAL CELL JUNCTIONS. THE SHORT ISOFORM IS FOUND BOTH IN
CC       ENDOTHELIAL CELLS AND THE HIGHLY SPECIALIZED EPITHELIAL JUNCTIONS
```

```
CC       OF RENAL GLOMERULI AND SERTOLI CELLS OF THE SEMINIFEROUS TUBULES.
CC   -!- PTM: Phosphorylated.
CC   -!- SIMILARITY: Belongs to the MAGUK family.
CC   -!- SIMILARITY: Contains 3 PDZ/DHR domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   -!- SIMILARITY: Contains 1 guanylate kinase-like domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L14837; AAA02891.1; -.
DR   PIR; A47747; A47747.
DR   HSSP; P31016; 1BFE.
DR   Genew; HGNC:11827; TJP1.
DR   MIM; 601009; -.
DR   GO; GO:0005913; C:cell-cell adherens junction; TAS.
DR   GO; GO:0005923; C:tight junction; TAS.
DR   GO; GO:0007043; P:intercellular junction assembly; TAS.
DR   InterPro; IPR008144; Guanylate_kin.
DR   InterPro; IPR008145; Guanylt/Ca.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR005417; ZonOcculdens.
DR   InterPro; IPR005418; ZonOccul S1.
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00625; Guanylate_kin; 1.
DR   Pfam; PF00595; PDZ; 3.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01597; ZONOCCLUDNS.
DR   PRINTS; PR01598; ZONOCCLUDNS1.
DR   SMART; SM00072; GuKc; 1.
DR   SMART; SM00228; PDZ; 3.
DR   SMART; SM00326; SH3; 1.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR   PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR   PROSITE; PS50106; PDZ; 3.
DR   PROSITE; PS50002; SH3; 1.
KW   Tight junction; SH3 domain; Alternative splicing; Repeat; Membrane;
KW   Phosphorylation.
FT   DOMAIN       11     98       PDZ 1.
FT   DOMAIN      174    252       PDZ 2.
FT   DOMAIN      409    490       PDZ 3.
FT   DOMAIN      504    572       SH3.
FT   DOMAIN      632    782       GUANYLATE KINASE.
FT   DOMAIN     1231   1236       POLY-PRO.
FT   DOMAIN     1414   1420       POLY-PRO.
FT   VARSPLIC    910    989       Missing (in isoform Short).
FT                                /FTId=VSP_003148.
SQ   SEQUENCE   1736 AA;  194721 MW;  508D01B7A0814FFE CRC64;

  Query Match             2.1%;  Score 158.5;  DB 1;  Length 1736;
  Best Local Similarity   19.9%;  Pred. No. 0.32;
  Matches  138;  Conservative   83;  Mismatches  239;  Indels  233;  Gaps   36;

Qy         99 DTLPWSKEPPRAATAGQQWALQTVLLEVLNRFMPPPN-NTPGRTFGRTLSGPSGLSRPTS 157
              |: |  :: |||:                : :  | |  :| ||    |  : |  : |
Db       1129 DSRPRYEQAPRASA-------------LRHEEQPAPGYDTHGRL--RPEAQPHPSAGPKP 1173

Qy        158 TNTKRKDEPANVTFAD-PPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLN 216
                :|:  |  : ::  ||:   :|:    |:||||    |  : :   :|    | |
Db       1174 AESKQYFEQYSRSYEQVPPQGFTSRAGHFEPLHGAA---AVPPLIPSSQHKP----EALP 1226

Qy        217 QCTKRAKGKLSDNVAAAAAPPVPIASALDKVP----------TRRHANTRDPTATGHRRA 266
                 ||               ||| |  :  :: |          :  | |  :    :
Db       1227 SNTK-------------PLPPPPTQTEEEEDPAMKPQSVLTRVKMFENKRSASLETKKDV 1273

Qy        267 DQVDSFDTSQGTS--YGSSVFSACRHNQSTTQSSFE---------APPSQPREKRPVDAT 315
                :   ||    :   |    |: :          :  |:|: |             | :|: | | |
Db       1274 NDTGSFKPPEVASKPSGAPIIGP----KPTSQNQFSEHDKTLYRIPEPQKPQLKPPED-- 1327

Qy        316 VFEAGHL--------------------IESPSKGRTTKSHIDNQPLSSSSQGETSFSTYY 355
               :  : |                     |:            ||:          :  ||  :::||:|
Db       1328 IVRSNHYDPEEDEEYYRKQLSYFDRRSFENKPPAHIAASHLSEPAKPAHSQNQSNFSSY- 1386

Qy        356 ESFPSSGGEGAIPEPSRSNGLARS-EESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPL 414
                  ||     :|  ::|:  ||        :|  |           |  |        :
Db       1387 ----SSKG-----KPPEADGVDRSFGEKRYEPIQATPPP---------PPLPSQYAQPSQ 1428

Qy        415 AVAWEVTRLFMHCK--------VDLE-DESLGLKYDPSWSTARDVTDIWKTLYRLDAFRG 465
                |        | :|  |          | |:     ||  |  ||  |              :
Db       1429 PVT--SASLHIHSKGAHGEGNSVSLDFQNSLVSKPDPPPS-----------------QN 1468

Qy        466 KPFPEKPPN--DVFVTAMTGNFESKGSAVVLSAVLDYNPDNSP----------TAPLYLV 513
              ||    :|||   |       ||      :    |                 || :|       | |
Db       1469 KPATFRPPNRED---TAQAAFYPQKSF-----------PDKAPVNGTEQTQKTVTPAY-N 1513

Qy        514 KLKPLMFEQGCR-LTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQ 572
                :  |  :     |    |:|   :|   |:||  |:   |  :      || |  : : :::
Db       1514 RFTPKPYTSSARPFERKFESPKFNHNLLPSETAHKPDLS---SKTPTSPKTLVK-----S 1565

Qy        573 HSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPK-------------PIIKERVH------ 613
              |||         |:    :    :    |  : ||  ||                |:      |
Db       1566 HSL---------AQPPEFDSGVETFSIHAEKPKYQINNISTVPKAIPVSPSAVEEDEDED 1616

Qy        614 -------------------FFAETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWL 654
                                   |||::      :    :   :|   |  |||    |||  :    |
Db       1617 GHTVVATARGIFNSNGGVLSSIETGVS----IIIPQGAIP--EGVEQEIYFKVCRDNSIL 1670

Qy        655 LQLDNNTWQ----------PH-LKLFSRIQLGL 676
                ||      :          || ||    ::| |
Db       1671 PPLDKEKGETLLSPLVMCGPHGLKFLKPVELRL 1703

RESULT 8
TAC2_HUMAN
ID   TAC2_HUMAN     STANDARD;      PRT;  1026 AA.
AC   O95359; Q9NZ41; Q9NZR5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transforming acidic coiled-coil-containing protein 2 (Anti Zuai-1)
DE   (AZU-1).
GN   TACC2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain, Fetal brain, and Skeletal muscle;
RX   MEDLINE=20570483; PubMed=11121038;
RA   Gergely F., Karlsson C., Still I.H., Cowell J.K., Kilmartin J.,
RA   Raff J.W.;
RT   "The TACC domain identifies a family of centrosomal proteins that can
RT   interact with microtubules.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:14352-14357(2000).
RN   [2]
RP   SEQUENCE OF 375-1026 FROM N.A. (ISOFORM 2).
RX   MEDLINE=20214826; PubMed=10749935;
RA   Chen H.-M., Schmeichel K.L., Mian I.S., Lelievre S., Petersen O.W.,
RA   Bissell M.J.;
RT   "AZU-1: a candidate breast tumor suppressor and biomarker for tumor
RT   progression.";
RL   Mol. Biol. Cell 11:1357-1367(2000).
RN   [3]
RP   SEQUENCE OF 44-1026 FROM N.A. (ISOFORM 3).
RA   Pu J., Li C., Rodriguez M., Banerjee D.;
RT   "Expression of TACC2 protein mRNA in human microvascular endothelial
RT   cells.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
```

```
CC   -!- FUNCTION: May play a role in organizing centrosomal microtubules.
CC       May act as a tumor suppressor protein. May represent a tumor
CC       progression marker.
CC   -!- SUBUNIT: Interacts with microtubules.
CC   -!- SUBCELLULAR LOCATION: Nuclear; concentrated at centrosomes.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC         Comment=Experimental confirmation may be lacking for some
CC         isoforms;
CC       Name=1;
CC         IsoId=O95359-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=O95359-2; Sequence=VSP_006368, VSP_006369;
CC       Name=3;
CC         IsoId=O95359-3; Sequence=VSP_006369;
CC   -!- SIMILARITY: Belongs to the TACC family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF095791; AAC64968.2; -.
DR   EMBL; AF176646; AAF63433.1; ALT_INIT.
DR   EMBL; AF220152; AAF29537.2; -.
DR   Genew; HGNC:11523; TACC2.
DR   MIM; 605302; -.
DR   InterPro; IPR007707; TACC.
DR   Pfam; PF05010; TACC; 1.
KW   Coiled coil; Nuclear protein; Alternative splicing.
FT   DOMAIN       43     51       POLY-PRO.
FT   DOMAIN      498    501       POLY-LYS.
FT   DOMAIN      753    781       COILED COIL (POTENTIAL).
FT   DOMAIN      824   1025       COILED COIL (POTENTIAL).
FT   VARSPLIC    507    510       Missing (in isoform 2).
FT                                /FTId=VSP_006368.
FT   VARSPLIC    711    787       Missing (in isoform 2 and isoform 3).
FT                                /FTId=VSP_006369.
FT   CONFLICT     44     67       PAPPPPPPEVIPEPEVSTQPPPEE -> MFWYKRRGAPMRP
FT                                VSVIDGVVCVS (IN REF. 3).
SQ   SEQUENCE   1026 AA;  112110 MW;  E2575FCB446E9CF8 CRC64;

  Query Match             2.1%;  Score 157;  DB 1;  Length 1026;
  Best Local Similarity   18.9%;  Pred. No. 0.17;
  Matches  227;  Conservative  163;  Mismatches  410;  Indels  398;  Gaps   53;

Qy         71 SLNQAEA-NFFIEAKAASSNWVPKAHADPDTLPWSKEPPRAATAGQQWALQTVLLEVLNR 129
Db          6 SLQPAPASDLNLEASEAMSSDSEEAFETPESTTPVKAPPAPP------------------ 47

Qy        130 FMPPPNNTPGRTFGRTLSGPSGLSRPTSTNTKRKDEPA----NVTFADPPKR-SLTRSAT 184
Db         48 --PPP-------------PEVIPEPEVSTQPPPEEPGCGSETVPVPDGPRSDSVEGSPF 91

Qy        185 GPPIHGAAIPLKFPDPVNTGSKRPSLESENLN------------QCTKRAKGKLSDNVAA 232
Db         92 RPPSHSFSAVFDEDKPI-ASSGTYNLDFDNIELVDTFQTLEPRASDAKNQEGKV--NTRR 148

Qy        233 AAAPPVPIA-SALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACR-- 289
Db        149 KSTDSVPISKSTLSRSLSLQASDFDGASSSGNPEAVALAPDAYSTGSSSASSTLKRTKKP 208

Qy        290 -----HNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKSH------IDN 338
Db        209 RPPSLKKKQTTKKPTETPPVKETQQEPDEESLVPSGENLASETKTESAKTEGPSPALLEE 268

Qy        339 QPLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARL 398
Db        269 TPLE----------------PAVGPKAACPLDSESAEGVVPPASGGGRVQNSPPVGRKTL 312

Qy        399 RNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLEDE-----SLGLKYD-----PSWSTAR 448
Db        313 -------------PLTTAPEAGEVTPSDSGGQEDSPAKGLSVRLEFDYSEDKSSWDNQQ 358

Qy        449 D-------------------VTDIWKTLYRLDAFRGKPFPEKP--PNDVFVTAMTGNFES 487
Db        359 ENPPPTKKIGKKPVAKMPLRRPKMKKTPEKLDNTPASP-PRSPAEPNDIPIAKGTYTFD- 416

Qy        488 KGSAVVLSAVLDY--NPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTS 545
Db        417 ----------IDKWDDPNFNPFSSTSKMQESPKLPQQ----SYNFDPDTCDESVDPFKTS 462

Qy        546 T-SPSVPPVVSKQPGAVEEVIQWLTMGQHSLVG-------------RQWRAFFAKDAGYR 591
Db        463 SKTPSSP---SKSPASFE--IPASAMEANGVDGDGLNKPAKKKKTPLKTDTFRVKKSPKR 517

Qy        592 KPLREFQLRAEDP--------KPIIKERVHFFAETGI----------------------- 620
Db        518 SPLSD--PPSQDPTPAATPETPPVISAVVHATDEEKLAVTNQKWTCMTVDLEADKQDYPQ 575

Qy        621 -----TFRPDVFKTRSVVPAEEPVEQRTEFKVSQM--------------LDWLLQLDNNT 661
Db        576 PSDLSTF---VNETKFSSPTEE-LDYRNSYEIEYMEKIGSSLPQDDDAPKKQALYLMFDT 631

Qy        662 WQPHLKLFSRIQLGLS----------KTYAIMT------------LEPHQIRHHKTDLLS 699
Db        632 SQESPVKSSPVRMSESPTPCSGSSFEETEALVNTAAKNQHPVPRGLAPNQESHLQVPEKS 691

Qy        700 PSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGM---WVIDVDDTGDE 756
Db        692 SQKELEAM--GLGTPSEAIE------------ITAPEGSFASADALLSRLAHPVSLCGAL 737

Qy        757 DWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAI 816
Db        738 DYLEP---------DLAEKNP-PLFAQKLQEELEFAIMRIEAL---------KLARQIAL 778

Qy        817 GDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLN 876
Db        779 ASRSHQDAKREAA-------HPTD----VSISKTALYSRIGTAEVEKPAGLLFQQPD--- 824

Qy        877 FLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVH 936
Db        825 ----------------------------LDSAL--QIARAEII---------TKEREV- 843

Qy        937 VGFSSKFRDE-EESFTLLSDCDVLVA---RSPAHFPSDIQRVRAVFKPELHSL----KDV 988
Db        844 ----SEWKDKYEESRREVMEMRKIVAEYEKTIAQMIEDEQREKSVSHQTVQQLVLEKEQA 899

Qy        989 IIFSTKGDVPLA------KKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYL-- 1040
Db        900 LADLNSVEKSLADLFRRYEKMK--------------EVLEGFRKNEEVLKRCAQEYLSR 944

Qy       1041 -KKDKTTFKQL------------------------MASHGTGSAAKEQTTYDMIQKS 1072
Db        945 VKKEEQRYQALKVHAEEKLDRANAEIAQVRGKAQQEQAAH-QASLRKEQLRVDALERT 1001

RESULT 9
NCO6_MOUSE
ID   NCO6_MOUSE     STANDARD;      PRT;   2067 AA.
AC   Q9JLI9; Q9JLT9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Nuclear receptor coactivator 6 (Amplified in breast cancer-3 protein)
DE   (Cancer-amplified transcriptional coactivator ASC-2) (Activating
DE   signal cointegrator-2) (ASC-2) (Peroxisome proliferator-activated
DE   receptor-interacting protein) (PPAR-interacting protein) (Nuclear
DE   receptor-activating protein, 250 kDa) (Nuclear receptor coactivator
DE   RAP250) (NRC) (Thyroid hormone receptor binding protein).
GN   NCOA6 OR AIB3 OR RAP250 OR PRIP OR TRBP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND INTERACTION WITH
RP   PPARA; PPARG; RARA; RXRA; ESR1; ESR2 AND THRB.
RC   TISSUE=Liver;
RX   MEDLINE=20250907; PubMed=10788465;
RA   Zhu Y.-J., Kan L., Qi C., Kanwar Y.S., Yeldandi A.V., Rao M.S.,
RA   Reddy J.K.;
RT   "Isolation and characterization of peroxisome proliferator-activated
RT   receptor (PPAR) interacting protein (PRIP) as a coactivator for
RT   PPAR.";
RL   J. Biol. Chem. 275:13510-13516(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Breast;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 786-1142 FROM N.A. (ISOFORM 1), INTERACTION WITH PPARA;
RP   PPARG; ESR1; ESR2; THRA AND THRB, AND MUTAGENESIS OF LEU-891 AND
RP   LEU-894.
RC   TISSUE=Embryo;
RX   MEDLINE=20148724; PubMed=10681503;
RA   Caira F., Antonson P., Pelto-Huikko M., Treuter E., Gustafsson J.-A.;
RT   "Cloning and characterization of RAP250, a nuclear receptor
RT   coactivator.";
RL   J. Biol. Chem. 275:5308-5317(2000).
RN   [4]
RP   INTERACTION WITH RNPC2.
RX   MEDLINE=21638469; PubMed=11704680;
RA   Jung D.-J., Na S.-Y., Na D.S., Lee J.W.;
RT   "Molecular cloning and characterization of CAPER, a novel coactivator
RT   of activating protein-1 and estrogen receptors.";
RL   J. Biol. Chem. 277:1229-1234(2002).
CC   -!- FUNCTION: Nuclear receptor coactivator that directly binds nuclear
CC       receptors and stimulates the transcriptional activities in a
CC       hormone-dependent fashion. Coactivates expression in an agonist-
CC       and AF2-dependent manner. Involved in the coactivation of
CC       different nuclear receptors, such as for steroids (GR and ERs),
CC       retinoids (RARs and RXRs), thyroid hormone (TRs), vitamin D3 (VDR)
CC       and prostanoids (PPARs). Probably functions as a general
CC       coactivator, rather than just a nuclear receptor coactivator. May
CC       also be involved in the coactivation of the NF-kappa-B pathway.
CC       May coactivate expression via a remodeling of chromatin and its
CC       interaction with histone acetyltransferase proteins. Involved in
CC       placental, cardiac, hepatic and embryonic development.
CC   -!- SUBUNIT: Monomer and homodimer. Interacts in vitro with the basal
CC       transcription factors GTF2A and TBP, suggesting an autonomous
CC       transactivation function. Interacts with NCOA1, CRSP3, RBM14, the
CC       histone acetyltransferase proteins EP300 and CREBBP, and with
CC       methyltransferase proteins NCOA6IP and HRMT1L1 (By similarity).
CC       Interacts with RNPC2. Belongs to the ASC-2/NCOA6 complex (ASCOM),
CC       which contains ASC-2/NCOA6, the retinoblastoma-binding protein
CC       RBQ-3/ RBBP5, alpha- and beta-tubulins, the trithorax group
CC       proteins MLL2 and MLL3, and ASH2/ASCL2 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9JL19-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9JL19-2; Sequence=VSP_003410;
CC         Note=Acts as a dominant negative repressor;
CC   -!- TISSUE SPECIFICITY: Widely expressed. High expression in testis
CC       and weak expression in small intestine.
CC   -!- DEVELOPMENTAL STAGE: Expressed at E9 in placenta and at weaker
CC       level in uterus. High expression in neural tube and in CNS
CC       throughout development. High expression in sensory ganglia and
CC       retina from E11. In the alimentary tract and olfactory epithelium
CC       expression was seen from E13. Strong expression present in liver
CC       and kidney, from E11 and E13 respectively, and then expression
CC       decreased at later stages of development. Moderate expression in
CC       lung from E13, while it decreases during postnatal life. Strong
CC       expression in thymus from E15 onwards, and in spleen from E17 and
CC       during early postnatal life, then, the expression decreases.
CC   -!- DOMAIN: Contains two Leu-Xaa-Xaa-Leu-Leu (LXXLL) motifs. Only
CC       motif 1 is essential for the association with nuclear receptors.
CC   -!- PTM: Phosphorylated (By similarity).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF216186; AAF35860.1; -.
DR   EMBL; BC031113; AAH31113.1; -.
DR   EMBL; AF135169; AAF35973.1; -.
DR   MGD; MGI:1929915; Ncoa6.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005667; C:transcription factor complex; IDA.
DR   GO; GO:0003682; F:chromatin binding; IDA.
DR   GO; GO:0030331; F:estrogen receptor binding; ISS.
DR   GO; GO:0046965; F:retinoid X receptor binding; ISS.
DR   GO; GO:0046966; F:thyroid hormone receptor binding; ISS.
DR   GO; GO:0003713; F:transcription co-activator activity; ISS.
DR   GO; GO:0016563; F:transcriptional activator activity; IDA.
DR   GO; GO:0007420; P:brain development; IMP.
DR   GO; GO:0001701; P:embryonic development (sensu Mammalia); IMP.
DR   GO; GO:0007507; P:heart development; IMP.
DR   GO; GO:0030099; P:myeloid blood cell differentiation; ISS.
DR   GO; GO:0006367; P:transcription initiation from Pol II promoter; IDA.
KW   Transcription regulation; Activator; Nuclear protein; Repeat;
KW   Alternative splicing.
FT   DOMAIN        1   1060       CREBBP-BINDING REGION (BY SIMILARITY).
FT   DOMAIN        1    932       TBP/GTF2A-BINDING REGION (BY SIMILARITY).
FT   DOMAIN        1   1314       NCOA1-BINDING REGION (BY SIMILARITY).
FT   DOMAIN      777    931       NCOA6IP-BINDING REGION (BY SIMILARITY).
FT   DOMAIN     1644   2067       EP300/CRSP3-BINDING REGION
FT                                (BY SIMILARITY).
FT   DOMAIN      227   1044       GLN-RICH.
FT   DOMAIN      376    381       POLY-PRO.
FT   DOMAIN      917    922       POLY-LYS.
FT   DOMAIN     1543   1592       SER-RICH.
FT   SITE        891    895       LXXLL MOTIF 1.
FT   SITE       1495   1499       LXXLL MOTIF 2.
FT   VARSPLIC    458   2067       Missing (in isoform 2).
FT                                /FTId=VSP_003410.
FT   MUTAGEN     891    894       LVNL->AVNA: ABOLISHES INTERACTION WITH
FT                                NUCLEAR RECEPTORS.
FT   CONFLICT     39     39       G -> S (IN REF. 2).
FT   CONFLICT    109    109       W -> R (IN REF. 2).
FT   CONFLICT    194    194       M -> I (IN REF. 2).
FT   CONFLICT    290    290       Q -> QQ (IN REF. 2).
```

```
FT   CONFLICT   1014   1014        P -> L (IN REF. 3).
FT   CONFLICT   1141   1142        SE -> RS (IN REF. 3).
SQ   SEQUENCE   2067 AA;  219663 MW;  C855F8777167AD48 CRC64;

  Query Match              2.1%;  Score 156;  DB 1;  Length 2067;
  Best Local Similarity   20.8%;  Pred. No. 0.58;
  Matches  130;  Conservative   74;  Mismatches  246;  Indels  176;  Gaps   30;

Qy          3 PITPRKRNSPVEEIINRLNNDYNLGLQCVADTTLTPHRRKELAESDEDFGRHDKIYRALN 62
Db       1063 PLNPDSQRMPVQQ---SGNVPVMVGLQGPASVPPSPDKQRMPMSVNTPMGSNSR-----K 1114

Qy         63 FLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLP------------WSKEP--- 107
Db       1115 MVYQENPQNSSSSPLGEMSSLPEASGSEVPSVAGGPNNMPSHLVVSQNQLMMTGPKPGPS 1174

Qy        108 PRAATAG---QQWALQTVLLEVLNRFMPPPNNTP---GRTFGRTLSGPSGLSRPTSTNTK 161
Db       1175 PLSATQGATPQQ---------------PPVNSLPSSHGHHF-PNVAAPTQTSRPKTPNRA 1218

Qy        162 R-----KDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTG-SKRPSL----E 211
Db       1219 SPRPYYPQTPNN----RPPSTEPSEISLSPERLNASIAGLFPPQINIPLPPRPNLNRGFD 1274

Qy        212 SENLNQCTKRAKGKLSDNVAAAAAP--PVPIASALDKVPTRRHANTRDPTATGHRRADQV 269
Db       1275 QQGLNPTTLKAIGQAPSNLTITNPPNFAAPQAHKLDSVVVNSGKQSNPGTT---KRASPS 1331

Qy        270 DSFDTSQGTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKRP-------VDATVFEAGHL 322
Db       1332 NSRRSSPGSS-----------RKTTPS-----PGRQNSKAPKLTLASQTSTTMLQN--- 1371

Qy        323 IESPSKGRTTKSHIDNQPLSSSSQGETSFSTYYESFPSSGG----EGAIPEPSRSNGLAR 378
Db       1372 MELPRNVLVGPTPLANPPLPG-------------SFPNNTGLNPQNPTVPVPAMGTVLED 1418

Qy        379 SEESA--------------RSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEVTRLF 424
Db       1419 NKESVNIPQDSDCQNAQGRKEQVNTELKVVPTQ-------------EAKMAVP------- 1458

Qy        425 MHCKVDLEDESL--GLKYDPS-WSTARDVTDIWKTLYR---------LDAFRGKPFPEKP 472
Db       1459 -------EDQSKKDGQPLDPNKLPSVEENKNLMSPAMREAPTSLSQLLDNSGAPNVTIKP 1511

Qy        473 P--NDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRF 530
Db       1512 PGLTDLEVTPPVVSGEDLRKASVIPTLQD-PPSKEPSTSL----SSPHSSEPCSTLAR-- 1564

Qy        531 GPDRFFEILIPSPTSTSPSVPPVVSK 556
Db       1565 ------SELSEVSSNAAPSIPPVMSR 1584

RESULT 10
PCLO_CHICK
ID   PCLO_CHICK     STANDARD;      PRT;  5120 AA.
AC   Q9PU36;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Piccolo protein (Aczonin) (Fragment).
GN   PCLO OR ACZ.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with rim and bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
CC   -!- FUNCTION: May act as a scaffolding protein involved in the
CC       organization of synaptic active zones and in synaptic vesicle
CC       trafficking (By similarity).
CC   -!- SUBUNIT: Interacts with Rabac1/Pra1 and profilin (By similarity).
CC   -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC       synaptic junctions (By similarity).
CC   -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC       phospholipids. Calcium binds with low affinity but with high
CC       specificity and induces a large conformational change.
CC   -!- SIMILARITY: Contains 2 C2 domains.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y19187; CAB60725.1; -.
DR   HSSP; P04410; 1A25.
DR   GO; GO:0045202; C:synaptic junction; ISS.
DR   GO; GO:0005509; F:calcium ion binding; ISS.
DR   GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
DR   GO; GO:0005522; F:profilin binding; ISS.
DR   GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR   GO; GO:0016080; P:synaptic vesicle targeting; ISS.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001565; Synaptotagmin.
DR   InterPro; IPR008899; Znf_piccolo.
DR   Pfam; PF00168; C2; 2.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF05715; Zf_piccolo; 2.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
DR   SMART; SM00239; C2; 2.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
DR   PROSITE; PS50106; PDZ; 1.
KW   Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
KW   Repeat.
FT   NON_TER       1      1
FT   DOMAIN      258    357       10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT                                P-A-K-P-Q-P-Q-Q-P-X.
FT   ZN_FING     368    392       C4-TYPE (POTENTIAL).
FT   ZN_FING     836    859       C4-TYPE (POTENTIAL).
FT   DOMAIN     2324   2343       POLY-PRO.
FT   DOMAIN     4414   4493       PDZ.
FT   DOMAIN     4627   4726       C2 DOMAIN 1.
FT   DOMAIN     5003   5094       C2 DOMAIN 2.
SQ   SEQUENCE   5120 AA;  560751 MW;  A658D9891B65B412 CRC64;

  Query Match              2.1%;  Score 154.5;  DB 1;  Length 5120;
  Best Local Similarity   18.7%;  Pred. No. 2.9;
  Matches  267;  Conservative  201;  Mismatches  524;  Indels  435;  Gaps   73;

Qy         99 DTLPWSKEP--------PRAATAGQQWALQTVLLEVLNRFMPP---------PNNTPGRT 141
Db        542 DTVPTSAAPGVKQDLADPQSPST-QQKVTDSPMPETTK---PPADTHPAGDKPDSKPLPQ 597

Qy        142 FGRTLSGPSGLSR---PTSTNTKRKDEPANVTFADPPKRSLTRSATGPPIHGAAIPLKFP 198
Db        598 VSRQKSDPKLASQSGAKSDAKTQKPSEPAPV--KDDPKKLQTKPAPKPDTKPAP---KGP 652

Qy        199 DPVNTG----SKRPSLESENLNQCTKRAKGKLSDNVAAAAAPPVPIASALDKVPTRRHAN 254
Db        653 Q-AGTGPRPTSAQPAPQPQQ-PQKTPEQSRRFSLNLGGITDAPKP-------QPT----- 698
```

```
Qy    255 TRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACRHNQST-----TQSSFEAPPS----Q 305
Db    699 TPQETVTG-------------KLFGFGASIFSQASSLISTAGQPGSQTSGPAPPATKQPQ 745

Qy    306 PREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQPLSSSSQGE-------TSFSTYYESF 358
Db    746 PPSQPPASQAPPKEAAQAQPPPKAAPTKK--ETKPLASEKLGPMASDSTLTTKGSDLEKK 803

Qy    359 PSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAW 418
Db    804 PSLAKDSKHQTAEAKKPAELSEQEKASQPKVSCPL------------------------- 838

Qy    419 EVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPP----- 473
Db    839 --------CKTGL---NIGSKDPPNFNTC---TECKKVVCNLCGFNPMPHIVEVQEWLCL 884

Qy    474 NDVFVTAMTGNFESKGSAVVLSAVLDYNPDNSPTA-----PLYLVKLKPLMFEQGCRLTR 528
Db    885 NCQTQRAMSGQLGDMGK-VPLPKLGPSQPVSKPPATPQKQPVPAVSHSP---------QK 934

Qy    529 RFGPDRFFEILIPSPTST----SPSVPPVVSK-QPGAVEEVIQWLTMGQHSLVGRQWRAF 583
Db    935 SSTP--------PTPAATKPKEEPSVPKEVPKLQQGKLEKTLSADKIQQ----------- 975

Qy    584 FAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFR-PDVFKTRSVVPAEEPVEQR 642
Db    976 -----GIQK---------EDAK----------SKQGKLFKTPSADKIQRVSQKEDSRLQQ 1011

Qy    643 TEFKVSQMLDWLL---QLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEP--HQIRHHK--- 694
Db   1012 TKLTKTPSSDKILHGVQKEDIKFQ-EAKL-AKIPSADKILHRLQKEDPKLQQMKMAKALS 1069

Qy    695 TDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWV--IDVDD 752
Db   1070 ADKIQPEAQKEDVQLQEVRLSKAV--------------SADKIQHGIQKDLNLQHVKIEK 1115

Qy    753 TGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNL----------QLLPVL 802
Db   1116 TSSVEKIQEAQKESKLQ---QDKLPKTLSEDKIPATVSSDHKKLLSKSEEDKKPELLEKS 1172

Qy    803 EDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVP 862
Db   1173 TPHPKDK-------KEQITAETTGHITEQKVEVEAPCD---------------------- 1203

Qy    863 FLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQK-RKCDTLKSKLNI-RVGRSAYI 920
Db   1204 ---KLHEKKQE----------DVKKEDLTTGIPQMVSKPEKAEEEKTPVPVSRLPRSDHV 1250

Qy    921 YMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSD-------CDVLVARSPAHFPSD--- 970
Db   1251 EAVRE---KIEKEDDKSDTSSS--QQQKSPQGLSDTGYSSDGISSSLGEIPSHIPSDEKD 1305

Qy    971 ----------IQRVRAVFKPELHSLKDVII------FSTKGDVPLAKKLSGGDYDGDMAW 1014
Db   1306 LPREPSQKDTISQESPPSPSDLAKLESTVLSILEAQASTLTDEKSVKRKE--------LY 1357

Qy   1015 VCWDPEIVDGFVNAEMPLEP--------DLSRYLKKDKTTFKQLMASHGTGSAAKEQTTY 1066
Db   1358 ETYSEQTKDQHKTKPLPVTPESYSSDEEDLEAIQEGERT-----IAADSKGGASSQ---- 1408

Qy   1067 DMIQKSFHFALQPNFLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNE 1126
Db   1409 ------------------TDYKE-----EDGGNDTPA--RRQRYDSVEDSSESE---NS 1439

Qy   1127 ASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAIDKEL------E 1180
Db   1440 PVPRRKRRASVGSSSS---DEYKRDDSQGSGDEE---DFIRKQIIEMSADEDASGSEDDE 1493

Qy   1181 AFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNR------IGEV 1234
Db   1494 FIRNQLKEISVTESQKK-------------EEVKSKAKG-----TVGKHRRMARKSSAGY 1535

Qy   1235 EKEYGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDKSGANYDSKVIRLLELSFLA 1294
Db   1536 DEDAGR---RHSWHDDDD-------ETFDE----SPEPKYRETKSQDGE-----ELAISG 1576

Qy   1295 DREMNTWALLRASTAFKLYYHKSPKFVWQMAGRQLAYIKAQ-------MTSRP-----GE 1342
Db   1577 GGGLRRFKTIELNSTITSKYSETPE---QQKG--ILYFDEEPELEMESLTDSPEDRSRGE 1631

Qy   1343 GAPALMTAFMYAGLMPDKKFTKQYVARLEGDGSEYPDPEVYEVLGDD 1389
Db   1632 GSSSLHASSFTPGTSPTS------VSSLDEDSDSSPS---HKKLGGE 1669
```

```
RESULT 11
MACF_HUMAN
ID   MACF_HUMAN     STANDARD;      PRT;  5430 AA.
AC   Q9UPN3; O75053; Q8WXY2; Q9H540; Q9UKP0; Q9ULG9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Microtubule-actin crosslinking factor 1, isoforms 1/2/3 (Actin cross-
DE   linking family protein 7) (Macrophin 1) (Trabeculin-alpha) (620 kDa
DE   actin-binding protein) (ABP620).
GN   MACF1 OR ACF7 OR ABP620 OR KIAA0465 OR KIAA1251.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RX   MEDLINE=20001959; PubMed=10529403;
RA   Okuda T., Matsuda S., Nakatsugawa S., Ichigotani Y., Iwahashi N.,
RA   Takahashi M., Ishigaki T., Hamaguchi M.;
RT   "Molecular cloning of macrophin, a human homologue of Drosophila
RT   kakapo with a close structural similarity to plectin and dystrophin.";
RL   Biochem. Biophys. Res. Commun. 264:568-574(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=20026884; PubMed=10559237;
RA   Sun Y., Zhang J., Kraeft S.-K., Auclair D., Chang M.-S., Liu Y.,
RA   Sutherland R., Salgia R., Griffin J.D., Ferland L.H., Chen L.B.;
RT   "Molecular cloning and characterization of human trabeculin-alpha, a
RT   giant protein defining a new family of actin-binding proteins.";
RL   J. Biol. Chem. 274:33522-33530(1999).
RN   [3]
RP   SEQUENCE OF 182-4812 FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=21833812; PubMed=11845288;
RA   Gong T.-W.L., Besirli C.G., Lomax M.I.;
RT   "MACF1 gene structure: a hybrid of plectin and dystrophin.";
RL   Mamm. Genome 12:852-861(2001).
RN   [4]
RP   SEQUENCE OF 868-2350 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20039619; PubMed=10574462;
RA   Nagase T., Ishikawa K.-I., Kikuno R., Hirosawa M., Nomura N.,
RA   Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:337-345(1999).
RN   [5]
RP   SEQUENCE OF 1544-5057 FROM N.A.
RA   Corby N.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE OF 3312-5430 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22158633; PubMed=12168954;
RA   Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT   "Construction of expression-ready cDNA clones for KIAA genes: manual
RT   curation of 330 KIAA cDNA clones.";
RL   DNA Res. 9:99-106(2002).
RN   [7]
```

```
RP   SEQUENCE OF 3734-5430 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98116662; PubMed=9455484;
RA   Seki N., Ohira M., Nagase T., Ishikawa K.-I., Miyajima N.,
RA   Nakajima D., Nomura N., Ohara O.;
RT   "Characterization of cDNA clones in size-fractionated cDNA libraries
RT   from human brain.";
RL   DNA Res. 4:345-349(1997).
CC   -!- FUNCTION: F-actin-binding protein which may play a role in cross-
CC       linking actin to other cytoskeletal proteins. Also binds to
CC       microtubules (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC       Name=2;
CC         IsoId=Q9UPN3-2; Sequence=Displayed;
CC       Name=1;
CC         IsoId=Q9UPN3-1; Sequence=VSP_007341;
CC       Name=3;
CC         IsoId=Q9UPN3-3; Sequence=Not described;
CC       Name=4;
CC         IsoId=Q96PK2-1; Sequence=External;
CC   -!- TISSUE SPECIFICITY: Ubiquitously expressed.
CC   -!- SIMILARITY: Belongs to the plakin or cytolinker family.
CC   -!- SIMILARITY: Contains 1 actin-binding domain.
CC   -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC   -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   -!- SIMILARITY: Contains 37 spectrin repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB029290; BAA83821.1; -.
DR   EMBL; AF141968; AAF06360.1; -.
DR   EMBL; AF325341; AAL39000.1; -.
DR   EMBL; AF325330; AAL39000.1; JOINED.
DR   EMBL; AF325331; AAL39000.1; JOINED.
DR   EMBL; AF325332; AAL39000.1; JOINED.
DR   EMBL; AF325333; AAL39000.1; JOINED.
DR   EMBL; AF325334; AAL39000.1; JOINED.
DR   EMBL; AF325335; AAL39000.1; JOINED.
DR   EMBL; AF325336; AAL39000.1; JOINED.
DR   EMBL; AF325339; AAL39000.1; JOINED.
DR   EMBL; AF325340; AAL39000.1; JOINED.
DR   EMBL; AB033077; BAA86565.1; -.
DR   EMBL; AL137853; CAC15920.1; -.
DR   EMBL; AB007934; BAA32310.2; -.
DR   PIR; T00079; T00079.
DR   HSSP; Q01082; 1BKR.
DR   Genew; HGNC:13664; MACF1.
DR   GO; GO:0005856; C:cytoskeleton; NAS.
DR   GO; GO:0003780; F:actin cross-linking activity; NAS.
DR   GO; GO:0005509; F:calcium ion binding; NAS.
DR   GO; GO:0008017; F:microtubule binding; NAS.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR001715; Calponin-like.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR003108; GAS2.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00036; efhand; 2.
DR   Pfam; PF02187; GAS2; 1.
DR   Pfam; PF00435; spectrin; 27.
DR   ProDom; PD000012; EF-hand; 1.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00054; EFh; 2.
DR   SMART; SM00243; GAS2; 1.
DR   SMART; SM00150; SPEC; 36.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR   PROSITE; PS50021; CH; 2.
DR   PROSITE; PS00018; EF_HAND; 2.
DR   PROSITE; PS50002; SH3; FALSE_NEG.
KW   Actin-binding; Cytoskeleton; Calcium; Calcium-binding; Repeat;
KW   SH3 domain; Coiled coil; Alternative splicing.
FT   DOMAIN        1    295       ACTIN-BINDING.
FT   DOMAIN       78    181       CH 1.
FT   DOMAIN      194    295       CH 2.
FT   DOMAIN      243    265       COILED COIL (POTENTIAL).
FT   DOMAIN      477    529       COILED COIL (POTENTIAL).
FT   DOMAIN      722    751       COILED COIL (POTENTIAL).
FT   DOMAIN      816    843       COILED COIL (POTENTIAL).
FT   DOMAIN     1013   1118       COILED COIL (POTENTIAL).
FT   DOMAIN     1164   1191       COILED COIL (POTENTIAL).
FT   DOMAIN     1399   1690       COILED COIL (POTENTIAL).
FT   DOMAIN     1780   1843       COILED COIL (POTENTIAL).
FT   DOMAIN     1975   2005       COILED COIL (POTENTIAL).
FT   DOMAIN     2039   2312       COILED COIL (POTENTIAL).
FT   DOMAIN     2385   2417       COILED COIL (POTENTIAL).
FT   DOMAIN     2544   2695       COILED COIL (POTENTIAL).
FT   DOMAIN     2760   2838       COILED COIL (POTENTIAL).
FT   DOMAIN     2911   3001       COILED COIL (POTENTIAL).
FT   DOMAIN     3130   3164       COILED COIL (POTENTIAL).
FT   DOMAIN     3244   3277       COILED COIL (POTENTIAL).
FT   DOMAIN     3418   3482       COILED COIL (POTENTIAL).
FT   DOMAIN     3596   3666       COILED COIL (POTENTIAL).
FT   DOMAIN     3786   3806       COILED COIL (POTENTIAL).
FT   DOMAIN     3852   3931       COILED COIL (POTENTIAL).
FT   DOMAIN     3967   3987       COILED COIL (POTENTIAL).
FT   DOMAIN     4084   4218       COILED COIL (POTENTIAL).
FT   DOMAIN     4343   4378       COILED COIL (POTENTIAL).
FT   DOMAIN     4408   4437       COILED COIL (POTENTIAL).
FT   DOMAIN     4468   4498       COILED COIL (POTENTIAL).
FT   DOMAIN     4907   4935       COILED COIL (POTENTIAL).
FT   DOMAIN     5044   5067       COILED COIL (POTENTIAL).
FT   REPEAT      314    355       SPECTRIN 1.
FT   REPEAT      591    623       SPECTRIN 2.
FT   REPEAT      680    784       SPECTRIN 3.
FT   REPEAT      786    800       SPECTRIN 4.
FT   DOMAIN      871    923       SH3.
FT   REPEAT     1250   1272       SPECTRIN 5.
FT   REPEAT     1287   1342       SPECTRIN 6.
FT   REPEAT     1455   1534       SPECTRIN 7.
FT   REPEAT     1547   1659       SPECTRIN 8.
FT   REPEAT     1815   1891       SPECTRIN 9.
FT   REPEAT     1932   2042       SPECTRIN 10.
FT   REPEAT     2260   2280       SPECTRIN 11.
FT   REPEAT     2372   2395       SPECTRIN 12.
FT   REPEAT     2398   2507       SPECTRIN 13.
FT   REPEAT     2510   2618       SPECTRIN 14.
FT   REPEAT     2621   2728       SPECTRIN 15.
FT   REPEAT     2731   2838       SPECTRIN 16.
FT   REPEAT     2841   2945       SPECTRIN 17.
FT   REPEAT     2987   3024       SPECTRIN 18.
FT   REPEAT     3136   3163       SPECTRIN 19.
FT   REPEAT     3187   3274       SPECTRIN 20.
FT   REPEAT     3277   3383       SPECTRIN 21.
FT   REPEAT     3386   3492       SPECTRIN 22.
FT   REPEAT     3495   3601       SPECTRIN 23.
FT   REPEAT     3604   3673       SPECTRIN 24.
FT   REPEAT     3713   3819       SPECTRIN 25.
FT   REPEAT     3832   3927       SPECTRIN 26.
FT   REPEAT     3982   4043       SPECTRIN 27.
FT   REPEAT     4046   4152       SPECTRIN 28.
FT   REPEAT     4155   4262       SPECTRIN 29.

  Query Match             2.1%;  Score 153;  DB 1;  Length 5430;
  Best Local Similarity   17.5%;  Pred. No. 3.9;
  Matches  217;  Conservative  175;  Mismatches  415;  Indels  436;  Gaps   58;
```

```
Qy    264 RRADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLI 323
Db   3765 RHKDSMDEL-----FSHRSEIFGTCGEEQKTVL----------QEKTESLIQQYEAISLL 3809

Qy    324 ESPSKGRTTKSHI-DNQPLSSSSQGETSFSTYYESFP----SSGGEGAIPEPSRSNGLAR 378
Db   3810 NSERYARLERAQVLVNQ---------FWETYEELSPWIEETRALIAQLPSPAIDHEQLR 3859

Qy    379 SEE----SARSQVQVHAP------VVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCK 428
Db   3860 QQQEEMRQLRESIAEHKPHIDKLLKIGPQLKELNPEEGEMVEEK----YQKAENMYAQIK 3915

Qy    429 VDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESK 488
Db   3916 EEVRQRALAL--DEAVSQSTQITE---------------FHDK------IEPMLETLENL 3952

Qy    489 GSAVVLSAVLDYNPD-------NSPTAPLYLVKLKPLMFEQGCR-----LTRRFGPDR-- 534
Db   3953 SSRLRMPPLIPAEVDKIRECISDNKSATVELEKLQP-SFEALKRRGEELIGRSQGADKDL 4011

Qy    535 --------------FFE---------------------------ILIPSPTST------ 546
Db   4012 AAKEIQDKLDQMVFFWEDIKARAEEREIKFLDVLELAEKFWYDMAALLTTIKDTQDIVHD 4071

Qy    547 --SPSVPPVVSKQPGAVEEVIQWLTMGQHS------LVGRQWRAFFA----KDAGYRKPL 594
Db   4072 LESPGIDPSIIKQQVEAAETIKEETDGLHEELEFIRILGAD--LIFACGETEKPEVRKSI 4129

Qy    595 REFQLRAEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFK--VSQMLD 652
Db   4130 DEMNNAWENLNKTWKERLE--------------------KLEDAMQAAVQYQDTLQAMFD 4169

Qy    653 WLLQLDNNTWQPHLKLFS-----------RIQLGLSKTYAIMTLEPHQIR----HHKTDL 697
Db   4170 W---LDNTV----IKLCTMPPVGTDLNTVKDQLNEMKEFKVEVYQ-QQIEMEKLNHQGEL 4221

Qy    698 LSPSGTGEVMNDGV-----------GRMSRSVAKRIRD----VLGLGDVPSAVQGRFGSA 742
Db   4222 MLKKATDETDRDIIREPLTELKHLWENLGEKIAHRQHKLEGALLALGQFQHALEELMS-- 4279

Qy    743 KGMWVIDVDD--------TGDEDWIETYPSQRKWECDFVDKHQRTLE-VRSVASEL--KS 791
Db   4280 ---WLTHTEELLDAQRPISGDPKVIEVELAKHHVLKNDVLAHQATVETVNKAGNELLESS 4336

Qy    792 AGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSS 851
Db   4337 AGDDASSL-------RSRLEAMNQCWESVLQKTEEREQQLQSTLQQAQGFHSEI-EDFLL 4388

Qy    852 RATRVS---HGRVPFLAGLPD----------------SQEETLNFLMN----------- 880
Db   4389 ELTRMESQLSASKP-TGGLPETAREQLDTHMELYSQLKAKEETYNQLLDKGRLMLLSRDD 4447

Qy    881 --SGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVG 938
Db   4448 SGSGSKTEQSVALLEQKWHVVSSKMEERKSKL-----------------------EEALN 4484

Qy    939 FSSKFRDEEESF---TLLSDCDVLVARSPAHFPSDIQ---RVRAVFKPELHSLKDVII-- 990
Db   4485 LATEFQNSLQEFINWLTLAEQSLNIASPPSLILNTVLSQIEEHKVFANEVNAHRDQIIEL 4544

Qy    991 ---------FSTKGDVPLAKKL-----------------SGGDYDG--------DMAWVC 1016
Db   4545 DQTGNQLKFLSQKQDVVLIKNLLVSVQSRWEKVVQRSIERGRSLDDARKRAKQFHEAW-- 4602

Qy   1017 WDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASH------------------GTGS 1058
Db   4603 --KKLIDWLEDAESHLDSELE--ISNDPDKIKLQLSKHKEFQKTLGGKQPVYDTTIRTGR 4658

Qy   1059 AAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQS 1118
Db   4659 ALKEKTLLPEDSQKL-----DNFLGEVRDKWDTVC-------------------GKSVERQ 4695

Qy   1119 ---KQGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAI 1175
Db   4696 HKLEEALLFS--------GQFMDALQALVDWLYKVEPQLAEDQPVH------------- 4733

Qy   1176 DKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEVE 1235
Db   4734 -GDLDLVMNLMDAHK-----------------VFQKE--------------LGKRTGTVQ 4761

Qy   1236 --KEYGR-LVKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDK 1275
Db   4762 VLKRSGRELIENS--RDDTTWVKGQLQELSTRWDTVCKLSVSK 4802
```

RESULT 12

```
MAC4_HUMAN
ID   MAC4_HUMAN     STANDARD;      PRT;  5938 AA.
AC   Q96PK2; Q8WXY1;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Microtubule-actin crosslinking factor 1, isoform 4.
GN   MACF1 OR ACF7 OR ABP620 OR KIAA0465 OR KIAA1251.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
RX   MEDLINE=21833812; PubMed=11845288;
RA   Gong T.-W.L., Besirli C.G., Lomax M.I.;
RT   "MACF1 gene structure: a hybrid of plectin and dystrophin.";
RL   Mamm. Genome 12:852-861(2001).
CC   -!- FUNCTION: May play a role in cross-linking cytoskeletal proteins
CC       by binding intermediate filaments to the N-terminal plectin
CC       repeats and microtubules to the C-terminus.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC       Name=4;
CC         IsoId=Q96PK2-1; Sequence=Displayed;
CC       Name=1;
CC         IsoId=Q9UPN3-1; Sequence=External;
CC       Name=2;
CC         IsoId=Q9UPN3-2; Sequence=External;
CC       Name=3;
CC         IsoId=Q9UPN3-3; Sequence=External;
CC   -!- TISSUE SPECIFICITY: Expressed in heart, lung, pituitary and
CC       placenta, not found in brain, kidney, liver, pancreas or skeletal
CC       muscle.
CC   -!- SIMILARITY: Belongs to the plakin or cytolinker family.
CC   -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC   -!- SIMILARITY: Contains 19 plectin repeats.
CC   -!- SIMILARITY: Contains 32 spectrin repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF317696; AAL09459.1; -.
DR   EMBL; AF325341; AAL38997.1; -.
DR   EMBL; AF325333; AAL38997.1; JOINED.
DR   EMBL; AF325334; AAL38997.1; JOINED.
DR   EMBL; AF325335; AAL38997.1; JOINED.
DR   EMBL; AF325336; AAL38997.1; JOINED.
DR   EMBL; AF325339; AAL38997.1; JOINED.
DR   EMBL; AF325340; AAL38997.1; JOINED.
DR   GO; GO:0005856; C:cytoskeleton; ISS.
DR   GO; GO:0005509; F:calcium ion binding; ISS.
DR   GO; GO:0008017; F:microtubule binding; ISS.
```

```
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR003108; GAS2.
DR   InterPro; IPR001101; Plectin_repeat.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00036; efhand; 2.
DR   Pfam; PF02187; GAS2; 1.
DR   Pfam; PF00681; Plectin; 11.
DR   Pfam; PF00435; spectrin; 26.
DR   ProDom; PD000012; EF-hand; 1.
DR   SMART; SM00054; EFh; 2.
DR   SMART; SM00243; GAS2; 1.
DR   SMART; SM00250; PLEC; 19.
DR   SMART; SM00150; SPEC; 32.
DR   PROSITE; PS00018; EF_HAND; 2.
KW   Cytoskeleton; Calcium; Calcium-binding; Repeat; Coiled coil;
KW   Alternative splicing.
FT   DOMAIN     1830   1936       COILED COIL (POTENTIAL).
FT   DOMAIN     2001   2192       COILED COIL (POTENTIAL).
FT   DOMAIN     2282   2345       COILED COIL (POTENTIAL).
FT   DOMAIN     2477   2507       COILED COIL (POTENTIAL).
FT   DOMAIN     2541   2654       COILED COIL (POTENTIAL).
FT   DOMAIN     2686   2814       COILED COIL (POTENTIAL).
FT   DOMAIN     2887   2919       COILED COIL (POTENTIAL).
FT   DOMAIN     3046   3197       COILED COIL (POTENTIAL).
FT   DOMAIN     3262   3503       COILED COIL (POTENTIAL).
FT   DOMAIN     3632   3666       COILED COIL (POTENTIAL).
FT   DOMAIN     3746   3779       COILED COIL (POTENTIAL).
FT   DOMAIN     3920   3984       COILED COIL (POTENTIAL).
FT   DOMAIN     4098   4168       COILED COIL (POTENTIAL).
FT   DOMAIN     4288   4308       COILED COIL (POTENTIAL).
FT   DOMAIN     4354   4386       COILED COIL (POTENTIAL).
FT   DOMAIN     4397   4433       COILED COIL (POTENTIAL).
FT   DOMAIN     4469   4489       COILED COIL (POTENTIAL).
FT   DOMAIN     4586   4720       COILED COIL (POTENTIAL).
FT   DOMAIN     4845   4880       COILED COIL (POTENTIAL).
FT   DOMAIN     4910   4939       COILED COIL (POTENTIAL).
FT   DOMAIN     4970   5000       COILED COIL (POTENTIAL).
FT   DOMAIN     5409   5437       COILED COIL (POTENTIAL).
FT   DOMAIN     5546   5569       COILED COIL (POTENTIAL).
FT   CA_BIND    5598   5610       EF-HAND 1 (POTENTIAL).
FT   CA_BIND    5634   5646       EF-HAND 2 (POTENTIAL).
FT   REPEAT       12     49       PLECTIN 1.
FT   REPEAT       53     88       PLECTIN 2.
FT   REPEAT       89    126       PLECTIN 3.
FT   REPEAT      130    164       PLECTIN 4.
FT   REPEAT      166    202       PLECTIN 5.
FT   REPEAT      203    240       PLECTIN 6.
FT   REPEAT      243    278       PLECTIN 7.
FT   REPEAT      279    316       PLECTIN 8.
FT   REPEAT      318    354       PLECTIN 9.
FT   REPEAT      725    762       PLECTIN 10.
FT   REPEAT      763    800       PLECTIN 11.
FT   REPEAT      801    838       PLECTIN 12.
FT   REPEAT      839    876       PLECTIN 13.
FT   REPEAT      897    934       PLECTIN 14.
FT   REPEAT      935    972       PLECTIN 15.
FT   REPEAT      975   1011       PLECTIN 16.
FT   REPEAT     1012   1049       PLECTIN 17.
FT   REPEAT     1121   1158       PLECTIN 18.
FT   REPEAT     1159   1196       PLECTIN 19.
FT   REPEAT     1925   2032       SPECTRIN 1.
FT   REPEAT     2052   2160       SPECTRIN 2.
FT   REPEAT     2211   2313       SPECTRIN 3.
FT   REPEAT     2320   2430       SPECTRIN 4.
FT   REPEAT     2437   2543       SPECTRIN 5.
FT   REPEAT     2550   2652       SPECTRIN 6.
FT   REPEAT     2659   2758       SPECTRIN 7.
FT   REPEAT     2765   2896       SPECTRIN 8.
FT   REPEAT     2903   3008       SPECTRIN 9.
FT   REPEAT     3015   3119       SPECTRIN 10.
FT   REPEAT     3126   3229       SPECTRIN 11.
FT   REPEAT     3236   3339       SPECTRIN 12.
FT   REPEAT     3346   3446       SPECTRIN 13.
FT   REPEAT     3453   3555       SPECTRIN 14.
FT   REPEAT     3562   3664       SPECTRIN 15.
FT   REPEAT     3671   3775       SPECTRIN 16.
FT   REPEAT     3782   3884       SPECTRIN 17.
FT   REPEAT     3891   3993       SPECTRIN 18.
FT   REPEAT     4000   4102       SPECTRIN 19.
FT   REPEAT     4109   4211       SPECTRIN 20.
FT   REPEAT     4218   4320       SPECTRIN 21.
FT   REPEAT     4327   4428       SPECTRIN 22.
FT   REPEAT     4438   4544       SPECTRIN 23.
FT   REPEAT     4551   4653       SPECTRIN 24.
FT   REPEAT     4660   4763       SPECTRIN 25.
FT   REPEAT     4770   4872       SPECTRIN 26.
FT   REPEAT     4879   4982       SPECTRIN 27.
FT   REPEAT     4989   5091       SPECTRIN 28.
FT   REPEAT     5098   5201       SPECTRIN 29.
FT   REPEAT     5208   5309       SPECTRIN 30.
FT   REPEAT     5316   5418       SPECTRIN 31.
FT   REPEAT     5425   5555       SPECTRIN 32.
FT   CONFLICT   1712   1712       P -> S (IN REF. 1; AAL38997).
SQ   SEQUENCE   5938 AA;  670134 MW;  B8784112752DA004 CRC64;

  Query Match             2.1%;  Score 153;  DB 1;  Length 5938;
  Best Local Similarity   17.6%;  Pred. No. 4.4;
  Matches  219;  Conservative  175;  Mismatches  412;  Indels  438;  Gaps   59;

Qy        264 RRADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSSFEAPPSQPREKRPVDATVFEAGHLI 323
              |  | :|            |: | :|  |   |            :||     :||  |:
Db       4267 RHKDSMDEL-----FSHRSEIFGTCGEEQKTVL----------QEKTESLIQQYEAISLL 4311

Qy        324 ESPSKGRTTKSHI-DNQPLSSSSQGETSFSTYYESFP----SSGGEGAIPEPSRSNGLAR 378
              |     |  :: :  ||                  : || |      :   :| |:  :   |
Db       4312 NSERYARLERAQVLVNQ----------FWETYEELSPWIEETRALIAQLPSPAIDHEQLR 4361

Qy        379 SEE----SARSQVQVHAP------VVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMHCK 428
               ::       |  :  | |         :  :|: :  |:  : : |      :  ::   |
Db       4362 QQQEEMRQLRESIAEHKPHIDKLLKIGPQLKELNPEEGEMVEEK----YQKAENMYAQIK 4417

Qy        429 VDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESK 488
               ::   :| |   |  :  |  :  :|:                  | :|      :  |    |:
Db       4418 EEVRQRALAL--DEAVSQSTQITE---------------FHDK------IEPMLETLENL 4454

Qy        489 GSAVVLSAVLDYNPD-------NSPTAPLYLVKLKPLMFEQGCR-----LTRRFGPDR-- 534
               | :  :  ::     |      : :|  : |  | ||:|  ||    |      :  |  | |:
Db       4455 SSRLRMPPLIPAEVDKIRECISDNKSATVELEKLQP-SFEALKRRGEELIGRSQGADKDL 4513

Qy        535 -------------FFE---------------------------ILIPSPTST------ 546
                           |:|                           |: :   |
Db       4514 AAKEIQDKLDQMVFFWEDIKARAEEREIKFLDVLELAEKFWYDMAALLTTIKDTQDIVHD 4573

Qy        547 --SPSVPPVVSKQPGAVEEVIQWLTMGQHS------LVGRQWRAFFA----KDAGYRKPL 594
                || :  | :  ||       | |:  |  |  |        ::|       ||       ||  :
Db       4574 LESPGIDPSIIKQQVEAAETIKEETDGLHEELEFIRILGAD--LIFACGETEKPEVRKSI 4631

Qy        595 REFQLRAEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFK--VSQMLD 652
                |       |:       |||:                          |: ::   :::   |  |
Db       4632 DEMNNAWENLNKTWKERLE-------------------KLEDAMQAAVQYQDTLQAMFD 4671

Qy        653 WLLQLDNNTWQPHLKLFS-----------RIQLGLSKTYAIMTLEPHQIR----HHKTDL 697
               |   |||       :||  :            : ||    | :  :    :  ||       :|: :|
Db       4672 W---LDNTV----IKLCTMPPVGTDLNTVKDQLNEMKEFKVEVYQ-QQIEMEKLNHQGEL 4723

Qy        698 LSPSGTGEVMNDGV-----------GRMSRSRVAKRIRD----VLGLGDVPSAVQGRFGSA 742
               :   |  |   | :                 :   :|  |          :|  ||     |::
Db       4724 MLKKATDETDRDIIREPLTELKHLWENLGEKIAHRQHKLEGALLALGQFQHALEELMS-- 4781

Qy        743 KGMWVIDVDD--------TGDEDWIETYPSQRKWECDFVDKHQRTLE-VRSVASEL--KS 791
                  |:   ::           :||   ||    ::        :  |   |||:|  |     :||    |
Db       4782 ---WLTHTEELLDAQRPISGDPKVIEVELAKHHVLKNDVLAHQATVETVNKAGNELLESS 4838

Qy        792 AGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSS 851
              || :      |     | :::      : ::   :  :  :  :   | :       |     :  |  :
Db       4839 AGDDASSL-------RSRLEAMNQCWESVLQKTEEREQQLQSTLQQAQGFHSEI-EDFLL 4890
```

```
Qy      852 RATRVS---HGRVPFLAGLPD-----------------SQEETLNFLMN----------- 880
Db     4891 ELTRMESQLSASKP-TGGLPETAREQLDTHMELYSQLKAKEETYNQLLDKGRLMLLSRDD 4949

Qy      881 --SGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVG 938
Db     4950 SGSGSKTEQSVALLEQKWHVVSSKMEERKSKL----------------------EEALN 4986

Qy      939 FSSKFRDEEESF---TLLSDCDVLVARSPAHFPSDIQ---RVRAVFKPELHSLKDVII-- 990
Db     4987 LATEFQNSLQEFINWLTLAEQSLNIASPPSLILNTVLSQIEEHKVFANEVNAHRDQIIEL 5046

Qy      991 ---------FSTKGDVPLAKKL-----------------SGGDYDG--------DMAWVC 1016
Db     5047 DQTGNQLKFLSQKQDVVLIKNLLVSVQSRWEKVVQRSIERGRSLDDARKRAKQFHEAW-- 5104

Qy     1017 WDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASH-----------------GTGS 1058
Db     5105 --KKLIDWLEDAESHLDSELE--ISNDPDKIKLQLSKHKEFQKTLGGKQPVYDTTIRTGR 5160

Qy     1059 AAKEQTTY-DMIQKSFHFALQPNFLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQ 1117
Db     5161 ALKEKTLLPEDTQK------LDNFLGEVRDKWDTVC-----------------GKSVER 5196

Qy     1118 S---KQGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPA 1174
Db     5197 QHKLEEALLFS--------GQFMDALQALVDWLYKVEPQLAEDQPVH------------ 5235

Qy     1175 IDKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEV 1234
Db     5236 --GDLDLVMNLMDAHK----------------VFQKE--------------LGKRTGTV 5262

Qy     1235 E--KEYGR-LVKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDK 1275
Db     5263 QVLKRSGRELIENS--RDDTTWVKGQLQELSTRWDTVCKLSVSK 5304
```

```
RESULT 13
TAC2_MOUSE
ID   TAC2_MOUSE     STANDARD;      PRT;   1035 AA.
AC   Q9JJG0;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Transforming acidic coiled-coil-containing protein 2.
GN   TACC2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Osada N., Kusuda J., Tanuma R., Ito A., Hirata M., Sugano S.,
RA   Hashimoto K.;
RT   "Isolation of full-length cDNA clones from mouse brain cDNA library
RT   made by oligo-capping method.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May play a role in organizing centrosomal microtubules.
CC       May act as a tumor suppressor protein (By similarity).
CC   -!- SUBUNIT: Interacts with microtubules (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear; concentrated at centrosomes (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the TACC family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB041546; BAA95031.1; -.
DR   HSSP; P02649; 1LE4.
DR   MGD; MGI:1928899; Tacc2.
DR   InterPro; IPR007707; TACC.
DR   Pfam; PF05010; TACC; 1.
KW   Coiled coil; Nuclear protein.
FT   DOMAIN       43     51       POLY-PRO.
FT   DOMAIN      496    499       POLY-LYS.
FT   DOMAIN      763    791       COILED COIL (POTENTIAL).
FT   DOMAIN      834   1034       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1035 AA;  112715 MW;  0698015282F94C78 CRC64;

  Query Match              2.0%;  Score 151;  DB 1;  Length 1035;
  Best Local Similarity   19.3%;  Pred. No. 0.4;
  Matches  203;  Conservative  127;  Mismatches  368;  Indels  354;  Gaps   51;
```

```
Qy       71 SLNQAEA-NFFIEAKAASSNWVPKAHADPDTLPWSKEPPRAATAGQQWALQTVLLEVLNR 129
Db        6 SLQPAPASDLNLEVSEAMSSDSEEAFETPESTTPVKAPPAPPP----------------- 48

Qy      130 FMPPPNNTPGRTFGRTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLTRSATGPPI- 188
Db       49 --PPPEVTP---------EPEVIDPPAPEEPGCISEPP-VVVPDGPRSS--ESVEGSPFR 94

Qy      189 --HGAAIPLKFPDPV-NTGSKRPSLES-------ENLNQCTKRAKGKLSDNVAAAAAPPV 238
Db       95 PSHSSSAVFDEDKPIASSGTYNLDFDSIELVDNFQSLEPCSADSKGQEC----------- 143

Qy      239 PIASALDKVPTRRHANTRDPTATG------HRRADQVDSFDTSQGTSYGSSVFSACRHNQ 292
Db      144 -------KVSTRRKSTESVPPSKSTLSRSLSLQASDFDGASCPGSPEAGTLTTDACGTGS 196

Qy      293 STTQSSFE-----APPS--------QPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQ 339
Db      197 NSASSTLKRTKKTRPPSLKKKQATKKPTETPPVKETQQEPGE--ESPV---PSEEHLAPE 251

Qy      340 PLSSSSQGETSFSTYYESFPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVA--AR 397
Db      252 TKTESATPEGAGCTLSDDTPLE--SPAVPTATCPLTLESAEDV--------SPLVSGGGR 301

Qy      398 LRNIWPKFPKWLHEAPLAVAWEVTRLFMHCKVDLED-----ESLGLKYD-----PSWSTA 447
Db      302 VQNSPPVGRK---SVPLTTASEAVEVTLSDSGGQEDLPAKGLSVRLEFDYSEDKGSWESQ 358

Qy      448 RDVT------------------DIWKTLYRLDAFRGKPFPEKP--PNDVFVTAMTGNFE 486
Db      359 QENAPPTKKIGKKPVAKMPLRRPKMKKTPEKLDNTPASP-PRSPTEPSDTPIAKGTYTFD 417

Qy      487 SKGSAVVLSAVLDY--NPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIP--- 541
Db      418 -----------IDKWDDPNFNPFSSTSKMQESPKLSQQ----SYNFDPDACEESLDPFKA 462

Qy      542 -SPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLR 600
Db      463 SSKTPSSPSKSPASFEIPASTTE--------------------ADGDGLNKPAK----K 497

Qy      601 AEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNN 660
Db      498 KKTPLKTMVEDVMSVCSLFDTFR--VKKSPKRSPLSDPPSQ------------------ 536

Qy      661 TWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAK 720
Db      537 ------------------------DP-----------TPAATPE-------------- 545

Qy      721 RIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVD------ 774
Db      546 ----------APSAIS------------TVVHATDEEKLAV--TSQKWTCMTVDLDADKQ 581

Qy      775 KHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHA 834
Db      582 DFPQPSDLSNFVNETKFNSPSEE----LDYRNSYEIEYMEKLGSSLPQD---DDTPKKQA 634
```

```
Qy        835 L--------NRPVE-----------------FRQWVYESYSSRATRVSHGRVPFLAGLPD 869
              |        ||:                 |      |:  : ||::  |   |   |||
Db        635 LYLMFDTPQESPVKSPPVRMSDSPTPCSGSSFED--TEALVNAATKLQH---PVARGLPS 689

Qy        870 SQEETLNFLMNSGFDPKK--QKYLQDIAWDLQKR------------KCDTLKSKLNIRVG 915
              |||  |          |:|  || |: :|                 ||| |:|
Db        690 SQEPLLQV-------PEKPSQKELEAMALGTPAEAIEITAPEGAFASADTLLSRL----- 737

Qy        916 RSAYIYMIADFWGVLEENEVHVG---FSSKFRDEEESFTLLSDCDVLVARSPA---HPPS 969
                |:   :      | || :          |: | ::|  | | ::     : :||  |
Db        738 --AHPASLCGALGYLEPDLAEKNPPVFAQKLQEELE-FAVMRIEALKLARQIALASRSRQ 794

Qy        970 DIQRVRAVFKPELHSLKDVIIFSTKGDVPLAK 1001
              | :|   |     |:: |:     ::|  |     : |
Db        795 DTKR-EAAHPPDV-SISKTALYSRIGSTEVEK 824
```

```
RESULT 14
MINT_MOUSE
ID   MINT_MOUSE     STANDARD;      PRT;  3644 AA.
AC   Q62504; Q80TN9; Q99PS4; Q9QZW2;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Msx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN   MINT OR SHARP OR KIAA0929.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE OF 49-3644 FROM N.A. (ISOFORM 1), FUNCTION, SUBCELLULAR
RP   LOCATION, TISSUE SPECIFICITY, DNA-BINDING, AND INTERACTION WITH MSX2.
RC   TISSUE=Testis;
RX   MEDLINE=99379811; PubMed=10451362;
RA   Newberry E.P., Latifi T., Towler D.A.;
RT   "The RRM domain of MINT, a novel msx2 binding protein, recognizes and
RT   regulates the rat osteocalcin promoter.";
RL   Biochemistry 38:10678-10690(1999).
RN   [2]
RP   SEQUENCE OF 1-112 FROM N.A.
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Fukuda S.,
RA   Hashizume W., Hayashida K., Hirozane T., Hori F., Imotani K.,
RA   Ishii Y., Itoh M., Kagawa I., Kawai J., Kojima Y., Kondo S., Konno H.,
RA   Koya S., Miyazaki A., Murata M., Nakamura M., Nomura K., Numazaki R.,
RA   Ohno M., Ohsato N., Saito R., Sakazume N., Sano H., Sasaki D.,
RA   Sato K., Shibata K., Shiraki T., Tagami M., Takeda Y., Waki K.,
RA   Watahiki A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 69-3644 FROM N.A. (ISOFORM 2), AND VARIANTS THR-348;
RP   PHE-762; PHE-773 AND LEU-933.
RC   STRAIN=ICR; TISSUE=Brain;
RA   Sakamoto T., Gotou T., Isagawa Y., Mimura H., Kimura K., Kawaichi M.;
RT   "MINT/spen negatively regulates Notch signaling by inhibiting RBP-
RT   J/Su(H) activity.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 318-578 FROM N.A.
RC   TISSUE=Cochlea;
RX   MEDLINE=97237053; PubMed=9119401;
RA   Crozet F., El-Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA   Hamel C., Fizames C., Levi-Acobas F., Depetris D., Mattei M.-G.,
RA   Weil D., Pujol R., Petit C.;
RT   "Cloning of the genes encoding two murine and human cochlear
RT   unconventional type I myosins.";
RL   Genomics 40:332-341(1997).
RN   [5]
RP   SEQUENCE OF 2598-3644 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22579291; PubMed=12693553;
RA   Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA   Nakajima D., Nagase T., Ohara O., Koga H.;
RT   "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT   II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT   cDNAs identified by screening of terminal sequences of cDNA clones
RT   randomly sampled from size-fractionated libraries.";
RL   DNA Res. 10:35-48(2003).
RN   [6]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=22261914; PubMed=12374742;
RA   Oswald F., Kostezka U., Astrahantseff K., Bourteele S., Dillinger K.,
RA   Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
RA   Schmid R.M.;
RT   "SHARP is a novel component of the Notch/RBP-Jkappa signalling
RT   pathway.";
RL   EMBO J. 21:5417-5426(2002).
RN   [7]
RP   FUNCTION, AND TISSUE SPECIFICITY.
RX   MEDLINE=22483652; PubMed=12594956;
RA   Kuroda K., Han H., Tani S., Tanigaki K., Tun T., Furukawa T.,
RA   Taniguchi Y., Kurooka H., Hamada Y., Toyokuni S., Honjo T.;
RT   "Regulation of marginal zone B cell development by MINT, a suppressor
RT   of Notch/RBP-J signaling pathway.";
RL   Immunity 18:301-312(2003).
CC   -!- FUNCTION: Essential corepressor protein, which probably regulates
CC       different key pathways such as the Notch pathway. Negative
CC       regulator of the Notch pathway via its interaction with RBPSUH,
CC       which prevents the association between NOTCH1 and RBPSUH, and
CC       therefore suppresses the transactivation activity of Notch
CC       signaling. Blocks the differentiation of precursor B cells into
CC       marginal zone B cells. Probably represses transcription via the
CC       recruitment of large complexes containing histone deacetylase
CC       proteins. May bind both to DNA and RNA.
CC   -!- SUBUNIT: Interacts with NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and
CC       MTA1L1. Interacts with the nuclear receptors RAR and PPARD.
CC       Interacts with RAR in absence of ligand. Bind to the steroid
CC       receptor RNA coactivator SRA (By similarity). Interacts with MSX2.
CC       Interacts with RBPSUH; this interaction may prevent the
CC       interaction between RBPSUH and NOTCH1.
CC   -!- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q62504-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q62504-2; Sequence=VSP_008564;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Highly expressed in testis. Expressed at lower
CC       level in brain, lung, spleen, liver and kidney. Weakly expressed
CC       in cardiac and skeletal muscles and ovary. In spleen, it is
CC       expressed in follicular B-cells, while it is weakly expressed in
CC       marginal zone B-cells.
CC   -!- DOMAIN: The RID domain mediates the interaction with nuclear
CC       receptors.
CC   -!- DOMAIN: The SPOC domain, which mediates the interaction with
CC       NCOR2, is essential for the repressive activity (By similarity).
CC   -!- SIMILARITY: Belongs to the Spen family.
CC   -!- SIMILARITY: Contains 1 RID (receptor interacting) domain.
CC   -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC   -!- SIMILARITY: Contains 1 SPOC domain.
CC   -!- CAUTION: Ref.4 sequence differs from that shown due to multiple
CC       frameshifts and conflicts that create stop codons.
CC   -!- CAUTION: Ref.5 sequence differs from that shown due to what seems
CC       to be the presence of intronic sequence in the cDNA.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; BY726481; -; NOT_ANNOTATED_CDS.
DR   EMBL; AF156529; AAD55931.1; ALT_INIT.
```

```
DR   EMBL; AB055980; BAB32786.1; -.
DR   EMBL; Z78160; CAB01562.1; ALT_SEQ.
DR   EMBL; AK122402; BAC65684.2; ALT_SEQ.
DR   MGD; MGI:1891706; Mint.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 3.
DR   SMART; SM00360; RRM; 3.
DR   PROSITE; PS50102; RRM; 4.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
DR   PROSITE; PS50917; SPOC; 1.
KW   Transcription regulation; Repressor; Nuclear protein; DNA-binding;
KW   RNA-binding; Repeat; Coiled coil; Alternative splicing; Polymorphism.
FT   DOMAIN        1    574       DNA-BINDING.
FT   DOMAIN        6     81       RNA-BINDING (RRM) 1.
FT   DOMAIN      336    416       RNA-BINDING (RRM) 2.
FT   DOMAIN      439    514       RNA-BINDING (RRM) 3.
FT   DOMAIN      518    590       RNA-BINDING (RRM) 4.
FT   DOMAIN      559    575       COILED COIL (POTENTIAL).
FT   DOMAIN      822    850       COILED COIL (POTENTIAL).
FT   DOMAIN     1185   1206       COILED COIL (POTENTIAL).
FT   DOMAIN     1509   1544       COILED COIL (POTENTIAL).
FT   DOMAIN     1607   1627       COILED COIL (POTENTIAL).
FT   DOMAIN     2216   2704       RID.
FT   DOMAIN     3478   3644       SPOC.
FT   DOMAIN     2138   2462       INTERACTION WITH MSX2.
FT   DOMAIN     2706   2845       INTERACTION WITH RBPSUH.
FT   DOMAIN      125    277       ARG-RICH.
FT   DOMAIN      236    326       SER-RICH.
FT   DOMAIN      648    721       TYR-RICH.
FT   DOMAIN      702    832       ARG-RICH.
FT   DOMAIN     2101   2233       ALA-RICH.
FT   DOMAIN     2377   2518       PRO-RICH.
FT   DOMAIN     2950   3475       PRO-RICH.
FT   VARSPLIC    618    640       Missing (in isoform 2).
FT                                /FTId=VSP_008564.
FT   VARIANT     348    348       I -> T.
FT   VARIANT     762    762       S -> F.
FT   VARIANT     773    773       S -> F.
FT   VARIANT     933    933       S -> L.
FT   CONFLICT    754    754       R -> G (IN REF. 3).
FT   CONFLICT   1524   1524       D -> A (IN REF. 3).
FT   CONFLICT   1560   1560       H -> Y (IN REF. 3).
FT   CONFLICT   1570   1570       F -> L (IN REF. 3).
FT   CONFLICT   1574   1574       R -> G (IN REF. 3).
FT   CONFLICT   1609   1609       Q -> R (IN REF. 3).
FT   CONFLICT   1659   1659       I -> V (IN REF. 3).
FT   CONFLICT   1669   1669       S -> F (IN REF. 3).
FT   CONFLICT   1705   1705       V -> A (IN REF. 3).
FT   CONFLICT   1815   1815       A -> V (IN REF. 3).
FT   CONFLICT   2097   2097       G -> A (IN REF. 3).
FT   CONFLICT   2201   2202       MISSING (IN REF. 3).
FT   CONFLICT   2322   2322       A -> V (IN REF. 3).
FT   CONFLICT   2385   2385       P -> Q (IN REF. 3).
FT   CONFLICT   2502   2502       R -> K (IN REF. 3).
FT   CONFLICT   2505   2505       E -> K (IN REF. 3).
FT   CONFLICT   2519   2519       D -> N (IN REF. 3).
FT   CONFLICT   2554   2554       T -> S (IN REF. 3).
FT   CONFLICT   2679   2688       LVSTPAGPVN -> VGEHPWWARD (IN REF. 3).
FT   CONFLICT   3010   3010       L -> P (IN REF. 3 AND 5).
SQ   SEQUENCE   3644 AA;  398750 MW;  9C7EC49A81A7DA4A CRC64;

  Query Match             2.0%;  Score 151;  DB 1;  Length 3644;
  Best Local Similarity   21.0%;  Pred. No. 2.8;
  Matches  137;  Conservative   63;  Mismatches  254;  Indels  198;  Gaps   29;

Qy          3 PITPRKRNSPVEEIINRLNNDYNLGLQCVADTTLTPHRRKELAESDEDFGRHDKIYRALN 62
                | |||:    |   :    |    |                 | |||  ||:
Db       1963 PTTPRRGRPP--KTRRRAEED-------------GEHERKEPAETPRPAEG--------- 1998

Qy         63 FLYWRKDDSLNQA----------------EANFFIEAKAASSNWVPKAHADPDTLPWSKE 106
                    ||   |    |               ||      |:|::     ||:  :  :     |
Db       1999 ---WRSPRSQKSAAAAGPQGKRGRNEQKVEAAAEAGAQASTREGNPKSRGEREAASEPKR 2055

Qy        107 PPRAATAGQQWALQTVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNTKRKDEP 166
                |   :   :       |   :|||  |     ||   |         ||   |   ||:  |          :
Db       2056 DRRDPSTDKS-GPDTFPVEVLER--KPPEKTYKSKRGRARSTRSGMDRAAHQRSLEMAAR 2112

Qy        167 ANVTFADPPKRSLTRSATGPPIHGAAIPLKFPDPVNTGSKRPSL-----------ESENL 215
                |     ||                 ||      || |  :     |          | |            | |:
Db       2113 AAGQAAD--------KEAGP---AAASPQESESPQKGSGSSPQLANNPADPDREAEEESA 2161

Qy        216 NQCTKRAKG-------KLSDNVAAAAAPPVPIASALDKVPTRRHANTRDPTAT------- 261
                :     |    :|          :|        |     |    |   | |:|   |          |    |||
Db       2162 SASTAPPEGTQLARQIELEQAVQNIAKLPEPSAAAASK-------GTATATATAASEEPA 2214

Qy        262 ---GHRRADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSSFEAPPS-------QPRE--- 308
                   ||:  |  |      :|        :  ||  :     |         |         :|  ||||            |||
Db       2215 PEHGHKPAHQAS--ETELAAAIGSIISDA-----SGEPENFSAPPSVPPGSQTHPREGME 2267

Qy        309 --KRPVDATVFEAGHLIESPSKGRTTKSHIDNQPLSSSSQGETSFSTYYESFPSSGGEGA 366
                        :: :  |  |     || :          |  :|    |    |  |:     ||    ::          :  |  :
Db       2268 PGLHEAESGILETGTATESSA---PQVSALD-PPEGSADTKETRGNSGDSVQEAKGSKAE 2323

Qy        367 IPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEVTRLFMH 426
                :    |  :    |     |  : :  |  :     :     |||                                  :|
Db       2324 VTPPRKDKG--RQKTTRRRKRNANKKVVA-----------------------ITETRAS 2357

Qy        427 CKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPND----VFVTAMT 482
                       :  ||          |:        |       |              :|   :  :    ||||:           |    :  |
Db       2358 EAEQTQSES------PAAEEATAATP--------EAPQEEKPSEKPPSPPAECTFDPSKT 2403

Qy        483 GNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEI--LI 540
                       ||                :: :    |   :|   |             :|  :  :                        ||     |::   :|
Db       2404 PPAESLSQE---NSAAEKTPCKAPVLPALPPLSQPALMDD--------GPQARFKVHSII 2452

Qy        541 PSPTSTSPS----------------VPPVVSKQPGAV------EEVIQWLT 569
                    |     |   ||                            |||:     ||   |                :|   :|:|
Db       2453 ESDPVTPPSDSGIPPPTIPLVTIAKLPPPVI---PGGVPHQSPPPKVTEWIT 2501

RESULT 15
SMC4_YEAST
ID   SMC4_YEAST     STANDARD;      PRT;  1418 AA.
AC   Q12267;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Structural maintenance of chromosome 4.
GN   SMC4 OR YLR086W OR L9449.5.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313267; PubMed=9169871;
RA   Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
RA   Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA   Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA   Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA   Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA   Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA   Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA   Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA   Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA   Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA   Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL   Nature 387:87-90(1997).
RN   [2]
RP   IDENTIFICIATION IN A CONDENSIN COMPLEX WITH SMC2; BRN1; YCS4 AND
RP   YCG1.
RX   MEDLINE=20273907; PubMed=10811823;
RA   Freeman L., Aragon-Alcaide L., Strunnikov A.;
RT   "The condensin complex governs chromosome condensation and mitotic
```

```
RT   transmission of rDNA.";
RL   J. Cell Biol. 149:811-824(2000).
CC   -!- FUNCTION: Central component of the condensin complex, a complex
CC       required for conversion of interphase chromatin into mitotic-like
CC       condense chromosomes. The condensin complex probably introduces
CC       positive supercoils into relaxed DNA in the presence of type I
CC       topoisomerases and converts nicked DNA into positive knotted forms
CC       in the presence of type II topoisomerases.
CC   -!- SUBUNIT: Forms an heterodimer with SMC2. Component of the
CC       condensin complex, which contains the SMC2 and SMC4 heterodimer,
CC       and three non SMC subunits that probably regulate the complex:
CC       BRN1, YCS4 and YCG1/YCS5.
CC   -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. In interphase
CC       cells, the majority of the condensin complex is found in the
CC       cytoplasm, while a minority of the complex is associated with
CC       chromatin. A subpopulation of the complex however remains
CC       associated with chromosome foci in interphase cells. During
CC       mitosis, most of the condensin complex is associated with the
CC       chromatin. At the onset of prophase, condensin associates with
CC       chromosome arms and to chromosome condensation. Dissociation from
CC       chromosomes is observed in late telophase.
CC   -!- DOMAIN: The flexible hinge domain, which separates the large
CC       intramolecular coiled coil regions, allows the heterodimerization
CC       with SMC2, forming a V-shaped heterodimer (By similarity).
CC   -!- SIMILARITY: Belongs to the SMC family. SMC4 subfamily.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z73258; CAA97646.1; -.
DR   EMBL; Z73259; CAA97648.1; -.
DR   EMBL; U53880; AAB67590.1; -.
DR   PIR; S64918; S64918.
DR   GermOnline; 142148; -.
DR   SGD; S0004076; SMC4.
DR   GO; GO:0000799; C:nuclear condensin complex; IPI.
DR   GO; GO:0007076; P:mitotic chromosome condensation; IMP.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR003405; SMC_C.
DR   InterPro; IPR003395; SMC_N.
DR   Pfam; PF02483; SMC_C; 1.
DR   Pfam; PF02463; SMC_N; 1.
KW   DNA condensation; Mitosis; Cell cycle; ATP-binding; Coiled coil;
KW   Nuclear protein.
FT   NP_BIND     185    192       ATP (POTENTIAL).
FT   DOMAIN      345    673       COILED COIL (POTENTIAL).
FT   DOMAIN      674    848       FLEXIBLE HINGE.
FT   DOMAIN      849   1172       COILED COIL (POTENTIAL).
FT   DOMAIN     1224   1263       COILED COIL (POTENTIAL).
FT   DOMAIN     1323   1358       ALA/ASP-RICH (DA-BOX).
SQ   SEQUENCE   1418 AA;  162188 MW;  F0E6B72FE8BFD374 CRC64;

  Query Match             2.0%;  Score 148.5;  DB 1;  Length 1418;
  Best Local Similarity   18.5%;  Pred. No. 0.92;
  Matches  209;  Conservative  174;  Mismatches  417;  Indels  329;  Gaps   53;

Qy        242 SALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACRHNQSTTQ---SS 298
              | | |      |:  |  ::|   :       || |: : ||           |   :::|  :     :
Db          4 SPLSKRQKRKSA--QEPELS----LDQGDAEEDSQ-------VENRVNLSENTPEPDLPA 50

Qy        299 FEAPPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDN-QPLSSSSQGETSFSTYYES 357
                || |:     | : :  |  :     |:  ||  |: |  ||  :||:|      :| :|
Db         51 LEASYSKSYTPRKLVLSSGENRYAFSQPTNSTTTSLHVPNLQPPKTSSRGR-DHKSYSQS 109

Qy        358 FPSSGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVA 417
               | | |        |:|   | :       |:|:         |: |: :      |::     :
Db        110 PPRSPGR----SPTRRLELLQLSPVKNSRVE---------LQKIYDR-----HQSS---S 148

Qy        418 WEVTRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVF 477
                : :|||::  |                              |        :: ||
Db        149 KQQSRLFINELV--------------------------LENFKSYAGK----------- 170

Qy        478 VTAMTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFE 537
                :  |  |  :   ||||        |: |  :  :     :  ::|  |  |      :  || :
Db        171 --QVVGPFHTSFSAVV-------GPNGSGKSNV----IDSMLFVFGFR-ANKMRQDRLSD 216

Qy        538 IL-----IPSPTSTSPSV--PPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGY 590
              ::        ||  |  | :|   |: :   |
Db        217 LIHKSEAFPSLQSCSVAVHFQYVIDESSGT----------------------------- 246

Qy        591 RKPLREFQLRAEDPKPIIKERVHFFAETGITFRPDVFKTRS----VVPAEEPVEQRTEFK 646
                       | :: ||               |:          ||   |     :   |     : |:
Db        247 --------SRIDEEKP-----------GLIITRKAFKNNSSKYYINEKESSYTEVTKLL 286

Qy        647 VSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEV 706
               :: :|     ||:        | |  :|     :     |  ::|        | :   |   |   |
Db        287 KNEGID----LDH-------KRFLILQ---GEVENIAQMKP------KAEKESDDGLLEY 326

Qy        707 MNDGVGRMSRS--VAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPS 764
               : | :|   :     : :|:  :   | :|       : ||          ::| :     |   ||
Db        327 LEDIIGTANYKPLIEERMGQIENLNEVCLEKENRFE------IVDREKNSLESGKET--- 377

Qy        765 QRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLE-------DRARDKVKMRQAIG 817
                    :|::|  ::       :||      :|             |         |    :|:|  ::::
Db        378 ----ALEFLEKEKQLTLLRSKLFQFKLLQSNSKLASTLEKISSSNKDLEDEKMKFQESL- 432

Qy        818 DRLINDLQRQFSEQKHALNR-PVEFRQWVYESYSSRATRVS-HGRVPFLAGLPDSQEETL 875
               : :::::  |   |  |  ::       :  :    |  |           ||||    |      |    :  |:||
Db        433 -KKVDEIKAQRKEIKDRISSCSSKEKTLVLERRELEGTRVSLEERTKNLVSKMEKAEKTL 491

Qy        876 NFLMNS------------GFDPKKQKYLQDIAWDLQKRK--CDTLKSKLNIRVGR-SAYI 920
                 :|              |    :  :  ::|:      |:|  :      | :|     |      :      || |
Db        492 KSTKHSISEAENMLEELRGQQTEHETEIKDLTQLLEKERSILDDIKLSLKDKTKNISAEI 551

Qy        921 Y---MIADFWGVL---EENEVHVGFSSKFRDEEESFTLLSDCDVL---VARSPAHFPSDI 971
                | :     :|::: :      |         ||     | :  :  |       :       ::
Db        552 IRHEKELEPWDLQLQEKESQIQLAESELSLLEETQAKLKKNVETLEEKILAKKTH-KQEL 610

Qy        972 QRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMP 1031
              | :      |  :|:||||            ||             :|             ::::          |
Db        611 QDLILDLKKKLNSLKDERSQGEKNFTSAHLKLK------EM------QKVLNAHRQRAME 658

Qy       1032 LEPDLSRYLKKDK--TTFKQLMAS------HG-----------------TGSAAKEQTTY 1066
                    ||:   |  |    |   :|   |          ||                         |        :
Db        659 ARSSLSKAQNKSKVLTALSRLQKSGRINGFHGRLGDLGVIDDSFDVAISTACPRLDDVVV 718

Qy       1067 DMIQKSFHFALQPNFLGMCTNY--KERLCYIN--------------NSVSNKP------ 1103
              | :: : |                 | :|  | :|  |                             ::  |  |
Db        719 DTVECAQH---------CIDYLRKNKLGYARFILLDRLRQFNLQPISTPENVPRLFDLV 768

Qy       1104 ------------AIILSSLVGNLVDQSKQ--------GIVFNEASWAQLRRELLGGALSL 1143
                              ::: :||    :  |:                     :|   :    :   :  ||   :
Db        769 KPKNPKFSNAFYSVLRDTLVAQNLKQANNVAYGKKRFRVVTVDGKLIDISGTMSGGGNHV 828

Qy       1144 PDPMYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDT-----EDGAHF 1198
                : |    ||  :  :   :  ||    :  :   |::||    |   :  |  ||      |:
Db        829 AKGLMK----LGTNQSDKVDDYTPEEVDK--IERELSERENNFRVASDTVHEMEEELKKL 882

Qy       1199 WD--PDLASYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNK 1245
                   |  |||  |   :   :   :|     |  |        |     :  ::  |  |    |  :  |  :|
Db        883 RDHEPDLESQISKAEMEAD---SLASELTLAEQQVKEAEMAYVKAVSDK 928

Search completed: March 29, 2004, 08:48:11
Job time : 27 secs
```

```
OM protein - protein search, using sw model

Run on:          March 29, 2004, 08:41:31 ; Search time 56 Seconds
                                            (without alignments)
                                            7899.224 Million cell updates/sec

Title:           US-09-913-878A-2
Perfect score:   7397
Sequence:        1 MNPITPRKRNSPVEEIINRL..........YEVLGDDDFDGIGFTGNGDY 1402

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:      1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_25:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10:  sp_plant:*
                 11:  sp_rodent:*
                 12:  sp_virus:*
                 13:  sp_vertebrate:*
                 14:  sp_unclassified:*
                 15:  sp_rvirus:*
                 16:  sp_bacteriap:*
                 17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7397	100.0	1402	3	Q9Y7G6	Q9y7g6 neurospora
2	328	4.4	1116	10	Q9ZRY7	Q9zry7 nicotiana t
3	328	4.4	1196	10	Q9SG02	Q9sg02 arabidopsis
4	326	4.4	1196	10	Q9LKP0	Q9lkp0 arabidopsis
5	325	4.4	1218	10	Q8LHH9	Q8lhh9 oryza sativ
6	322.5	4.4	775	10	Q9ZRX3	Q9zrx3 petunia hyb
7	320.5	4.3	1136	10	Q7XM31	Q7xm31 oryza sativ
8	312	4.2	1351	3	Q872R5	Q872r5 neurospora
9	298	4.0	1964	5	Q95ZG6	Q95zg6 dictyosteli
10	297	4.0	1114	10	Q9ZR58	Q9zr58 lycopersico
11	296.5	4.0	1780	5	Q19285	Q19285 caenorhabdi
12	295	4.0	1107	10	Q9LQV2	Q9lqv2 arabidopsis
13	294.5	4.0	1484	3	Q8TGV4	Q8tgv4 diaporthe p
14	293.5	4.0	1883	5	Q95ZG7	Q95zg7 dictyosteli
15	287	3.9	1107	10	Q8H1K9	Q8h1k9 arabidopsis
16	285.5	3.9	513	10	Q9ZRX5	Q9zrx5 arabidopsis
17	284	3.8	1638	3	Q9C162	Q9c162 neurospora
18	282.5	3.8	1122	3	Q8X1C2	Q8x1c2 phomopsis s
19	279.5	3.8	1133	10	O82504	O82504 arabidopsis
20	274	3.7	1222	10	Q94IV6	Q94iv6 oryza sativ
21	269.5	3.6	1483	3	Q96UL6	Q96ul6 diaporthe a
22	265.5	3.6	1215	3	O14227	O14227 schizosacch
23	263	3.6	1579	5	Q9BH56	Q9bh56 caenorhabdi
24	236.5	3.2	1069	5	Q9XYP1	Q9xyp1 dictyosteli
25	232.5	3.1	1601	5	Q9NDH1	Q9ndh1 caenorhabdi
26	230	3.1	1632	5	Q93593	Q93593 caenorhabdi
27	218.5	3.0	1377	5	Q95WU3	Q95wu3 giardia lam
28	217.5	2.9	2001	5	Q8MX71	Q8mx71 giardia lam
29	204.5	2.8	1148	10	Q94IV7	Q94iv7 oryza sativ
30	191	2.6	5105	5	O61201	O61201 caenorhabdi
31	177.5	2.4	966	10	O82190	O82190 arabidopsis
32	173.5	2.3	2291	5	Q9W1A9	Q9w1a9 drosophila
33	169.5	2.3	905	10	O82189	O82189 arabidopsis
34	161.5	2.2	1346	5	Q9U183	Q9u183 leishmania
35	159.5	2.2	1542	4	Q9Y6X0	Q9y6x0 homo sapien
36	159	2.1	929	10	O82188	O82188 arabidopsis
37	156	2.1	953	3	Q01869	Q01869 phanerochae
38	156	2.1	1116	11	Q8CHD2	Q8chd2 mus musculu
39	154.5	2.1	514	11	Q8R0K6	Q8r0k6 mus musculu
40	152.5	2.1	1769	6	O97758	O97758 canis famil
41	152.5	2.1	2271	3	Q9C443	Q9c443 cryptococcu
42	152.5	2.1	2948	4	Q86WG6	Q86wg6 homo sapien
43	152	2.1	1127	5	Q9V5B8	Q9v5b8 drosophila
44	151.5	2.0	807	4	Q8TCK9	Q8tck9 homo sapien
45	151	2.0	820	3	Q875A6	Q875a6 podospora a

ALIGNMENTS

```
RESULT 1
Q9Y7G6
ID   Q9Y7G6      PRELIMINARY;      PRT;  1402 AA.
AC   Q9Y7G6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   RNA-dependent RNA polymerase (Fragment).
GN   QDE-1.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cogoni C., Macino G.;
RT   "Gene silencing in neurospora crassa requires a protein homologous to
RT   RNA-dependent RNA polymerase.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ133528; CAB42634.1; -.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   InterPro; IPR007855; RdRP.
DR   Pfam; PF05183; RdRP; 1.
KW   RNA-directed RNA polymerase.
FT   NON_TER     1402    1402
SQ   SEQUENCE   1402 AA;  157983 MW;  FC45A1A17260837D CRC64;

  Query Match            100.0%;  Score 7397;  DB 3;  Length 1402;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1402;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNPITPRKRNSPVEEIINRLNNDYNLGLQCVADTTLTPHRRKELAESDEDFGRHDKIYRA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNPITPRKRNSPVEEIINRLNNDYNLGLQCVADTTLTPHRRKELAESDEDFGRHDKIYRA 60

Qy         61 LNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLPWSKEPPRAATAGQQWALQ 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LNFLYWRKDDSLNQAEANFFIEAKAASSNWVPKAHADPDTLPWSKEPPRAATAGQQWALQ 120
```

```
Qy        121 TVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TVLLEVLNRFMPPPNNTPGRTFGRTLSGPSGLSRPTSTNTKRKDEPANVTFADPPKRSLT 180

Qy        181 RSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQCTKRAKGKLSDNVAAAAAPPVPI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RSATGPPIHGAAIPLKFPDPVNTGSKRPSLESENLNQCTKRAKGKLSDNVAAAAAPPVPI 240

Qy        241 ASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSSFE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSSFE 300

Qy        301 APPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQPLSSSSQGETSFSTYYESFPS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 APPSQPREKRPVDATVFEAGHLIESPSKGRTTKSHIDNQPLSSSSQGETSFSTYYESFPS 360

Qy        361 SGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SGGEGAIPEPSRSNGLARSEESARSQVQVHAPVVAARLRNIWPKFPKWLHEAPLAVAWEV 420

Qy        421 TRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFVTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TRLFMHCKVDLEDESLGLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFVTA 480

Qy        481 MTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILI 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 MTGNFESKGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILI 540

Qy        541 PSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLR 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 PSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLR 600

Qy        601 AEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNN 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNN 660

Qy        661 TWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAK 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAK 720

Qy        721 RIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTL 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 RIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTL 780

Qy        781 EVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVE 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 EVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVE 840

Qy        841 FRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQK 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 FRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQK 900

Qy        901 RKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLV 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLV 960

Qy        961 ARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPE 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPE 1020

Qy       1021 IVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPN 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 IVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPN 1080

Qy       1081 FLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 FLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGA 1140

Qy       1141 LSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWD 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 LSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWD 1200

Qy       1201 PDLASYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQ 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PDLASYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDSKDPYPVRVNQ 1260

Qy       1261 VYEKWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHKSPKF 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 VYEKWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHKSPKF 1320

Qy       1321 VWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGDGSEYPDP 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 VWQMAGRQLAYIKAQMTSRPGEGAPALMTAFMYAGLMPDKKFTKQYVARLEGDGSEYPDP 1380

Qy       1381 EVYEVLGDDDFDGIGFTGNGDY 1402
              ||||||||||||||||||||||
Db       1381 EVYEVLGDDDFDGIGFTGNGDY 1402

RESULT 2
Q9ZRY7
ID   Q9ZRY7      PRELIMINARY;      PRT;  1116 AA.
AC   Q9ZRY7;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   RNA-directed RNA polymerase.
GN   RDRP.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Petit Havana SRI; TISSUE=Leaf;
RA   Schiebel W., Pelissier T., Riedel L., Thalmeir S., Schiebel R.,
RA   Kempe D., Lottspeich F., Saenger H.L., Wassenegger M.;
RT   "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from
RT   Tomato.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ011576; CAA09697.1; -.
DR   PIR; T30828; T30828.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   InterPro; IPR007855; RdRP.
DR   Pfam; PF05183; RdRP; 1.
KW   RNA-directed RNA polymerase.
SQ   SEQUENCE   1116 AA;  127553 MW;  8D9B89DA826D31DB CRC64;

  Query Match             4.4%;  Score 328;  DB 10;  Length 1116;
  Best Local Similarity   19.9%;  Pred. No. 2.9e-13;
  Matches  211;  Conservative  149;  Mismatches  386;  Indels  314;  Gaps   42;

Qy        441 DPSWSTARDVTDIWKTL-----------YRLDAFRGKPFPEKPPNDVFV--TAMTGNFES 487
              |  |    | |  |  |            :|  |    |  |   : |:  |  | :|
Db        193 DDQWVRTTDFTPSWIGLSSSLCLEFRNGVQLPNFSESFFYYKESMNQFILQTGFTFSFSQ 252

Qy        488 KGSAVVLSAVLDYNPDNSPTAPLYLVKLKPLMFEQGC--RLTRRFGPDRFFEILIPSPTS 545
              | : | :     : |:        | |:  |: : ||   |   |    ||::: |   :
Db        253 KLALVPIV----HLPEGIELPYKILFKISSLI-QHGCFPGLALNF---NFFQLVDPRRRN 304

Qy        546 TSPSVPPVVSKQPGAVEEV----------IQWLT-MGQHSLVGRQ--------------- 579
                      ::    |:|::          ::|||      | |||
Db        305 ---------IACIEHALEKLYYLKECCYDPVRWLTEQYDEYLKGRQLPKSPPITLDDGLVY 356

Qy        580 --------WRAFFAKDAGYRKPLREFQLR--AEDPKPIIK--------ERVHFF-----A 616
                       : :|    |    :    ||  :||    ::        |::|       |
Db        357 VRRVVVTPCKVYF---CGPEVNVSNRVLRNYSEDINNFLRVSFVDEEWEKIHSTDLLPRA 413

Qy        617 ETGITFRPDVFK-----TRSVVPAEEPVEQRTEFKVSQMLD---W-------LLQLDNNT 661
               ||   | |:::      |:     :   :   |  ||: |   |       |   |  |
Db        414 STGNGTRTDIYERILSTLRNGFIIGDKRFEFLAFSSSQLRDNSVWMFASRPGLTANDIRT 473

Qy        662 WQPHLKLFSRI-----QLGLSKTYAIMTLEPHQIRHHKTDLL-----SPSGTGEVMNDGV 711
              |    :    :     :|| |   :  ||    :  |: :::     |  ||  : :||:
Db        474 WMGDFRQIRNVAKYAARLGQSFGSSRETL---SVGRHEVEVIPDVACSLHGTNYIFSDGI 530
```

```
Qy      712 GRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECD 771
Db      531 GKISADFAHRVTIKCGLQYTPSSFQIRYGGYKGVVAVD-----------PYSSMK----- 574

Qy      772 FVDKHQRTLEVRSVASELKSAGLNLQLL------PVLEDRARDKVKMRQAIGDRLINDLQ 825
Db      575 --------LSLRKSMLKYESNNIKLDVLGWSKYQPCYLNRQLVTLLSTLGVIDDVLEQKQ 626

Qy      826 RQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDP 885
Db      627 NEAVDQLDAILHDSLKAQEALELMS-----------------PGENTNILKEMLNCGYMP 669

Qy      886 KKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSS---- 941
Db      670 DAEPFLSMMLQTFRASKLLDLRTRTRIFIPNGRTMMGCLDESRTLEYGQVFVQFSGAGRR 729

Qy      942 KFRDEEESF--TLLSDCD------VLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFST 993
Db      730 QFFEESHPFNDSGSANCDFILKGNVVVAKNPCLHPGDIRVLRAVDVPALHHMVDCVVFPQ 789

Qy      994 KGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMAS 1053
Db      790 KGKRPHPNECSGSDLDGDIYFVCWDPDLIPPRQVQSMDYTPAPTTQLDHD---------- 839

Qy     1054 HGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNSVS------------- 1100
Db      840 ----------VTIEEVEEYF-----------TNY-----IINDSLGIIANAHVVFADRE 872

Qy     1101 -----NKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPD-PMYKSDSWL 1154
Db      873 PDMAMSDPCKQLAQLFSIAVDFPKTGVPAEIPS--QLRPKEYPDFMEKPDKPTYPSEKVI 930

Qy     1155 GRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEIS 1214
Db      931 GK----------------------------LFQKVKNK-------TPQASSIATFTRDVA 955

Qy     1215 DKSRSSALLFTTLKNRIGEV---EKEYGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPE 1271
Db      956 RKSYDSDMIVDGFEDYIDEAFYYKSEY-----DNKLGNLMDYYGIKTE------AEILSG 1004

Qy     1272 AMDKSGANYDSK--------VIRLLELSFLADREMNTW---------ALLRASTAFKLYY 1314
Db     1005 GIMKASKTFDRRKDAEAIGVAVRCLR------KEARAWFKRRSDIDDMLAKASAWYHVTY 1058

Qy     1315 HKS--------------PKFVWQMAGRQLAYIKAQMTSRP 1340
Db     1059 HHTYWGLYNEGLKRDHFISFPWCVYDQLIQIKKAKARKRP 1098
```

RESULT 3
Q9SG02

```
ID   Q9SG02     PRELIMINARY;      PRT;  1196 AA.
AC   Q9SG02;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative RNA-directed RNA polymerase.
GN   T1G12.20 OR T9C5.95.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA   Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
RA   Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT   "Arabidopsis thaliana chromosome 1 BAC T1G12 genomic sequence.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,
RA   Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20306668; PubMed=10850495;
RA   Mourrain P., Beclin C., Elmayan T., Feuerbach F., Godon C.,
RA   Morel J.-B., Jouette D., Lacombe A.-M., Nikic S., Picault N.,
RA   Remoue K., Sanial M., Vo T.-A., Vaucheret H.;
RT   "Arabidopsis SGS2 and SGS3 Genes are Required for Posttranscriptional
RT   Gene Silencing and Natural Virus Resistance.";
RL   Cell 101:533-542(2000).
DR   EMBL; AC012329; AAG52184.1; -.
DR   EMBL; AL132964; CAB71285.1; -.
DR   EMBL; AF239718; AAF73959.1; -.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   InterPro; IPR007855; RdRP.
DR   Pfam; PF05183; RdRP; 1.
KW   RNA-directed RNA polymerase.
SQ   SEQUENCE   1196 AA;  136927 MW;  812DEB9CEDC82C8F CRC64;

  Query Match             4.4%;  Score 328;  DB 10;  Length 1196;
  Best Local Similarity   22.7%;  Pred. No. 3.2e-13;
  Matches  197;  Conservative  126;  Mismatches  312;  Indels  234;  Gaps   42;

Qy      535 FFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPL 594
Db      466 FVAPIVKDLTSSSFSQKTYVFKRVKSI------LTDG-FKLCGRKY-SFLAFSAN----- 512

Qy      595 REFQLRAEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWL 654
Db      513 ---QLR--------DRSAWFFAEDG--------KTR----------------VSDIKTWM 537

Qy      655 LQL-DNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGR 713
Db      538 GKFKDKNV----AKCAARMGLCFSSTYATVDVMPHEV---DTEVPDIERNGYVFSDGIGT 590

Qy      714 MSRSVAKRIRDVLGLGDV---PSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQ----- 765
Db      591 ITPDLADEVMEKLKL-DVHYSPCAYQIRYAGFKGV-------------VARWPSKSDGIR 636

Qy      766 ---RKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLIN 822
Db      637 LALRDSMKKFFSKH-TILEICS-WTRFQPGFLNRQIITLLS---------VLGVPDEIFW 685

Qy      823 DLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSG 882
Db      686 DMQ---ESMLYKLNRILDDTDVAFEVLTASCA---------------EQGNTAAIMLSAG 727

Qy      883 FDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFS-- 940
Db      728 FKPKTEPHLRGMLSSVRIAQLWGLREKSRIFVTSGRWLMGCLDEAGILEHGQCFIQVSKP 787

Qy      941 ----------SKFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVII 990
Db      788 SIENCFSKHGSRFKETKTDLEVVKGY-VAIAKNPCLHPGDVRILEAVDVPQLHHMYDCLI 846

Qy      991 FSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRY--LKKDKTTFK 1048
Db      847 FPQKGDRPHTNEASGSDLDGDLYFVAWDQKLI----------PPNRKSYPAMHYDAAEEK 896

Qy     1049 QLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTN----YKERLCYINNSVSNKPA 1104
Db      897 SL------GRAVNHQDIIDFFARN----LANEQLGTICNAHVVHADRSEY---GAMDEEC 943

Qy     1105 IILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPD-PMYKSDSWLGR------- 1156
Db      944 LLLAELAATAVDFPKTGKIVSMP--FHLKPKLYPDFMGKEDYQTYKSNKILGRLYRRVKE 1001

Qy     1157 -------------GEPTHI-----IDYLKFSIARP-------AIDKELEAFHNAMKAAKD 1191
```

```
Db       1002 VYDEDAEASSEESTDPSAIPYDAVLEIPGFEDLIPEAWGHKCLYDGQLIGLLGQYKVQKE 1061

Qy       1192 TE-DGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRLVKNKEMRDS 1250
Db       1062 EEIVTGHIWS---MPKYTSKKQGELKER--------LKHSYNSLKKEF-RKVFEETIPDH 1109

Qy       1251 KDPYPVRVNQVYEK----WCAIT--PEAMDKSGANYDSKVIRLLELSFLADREMNTWALL 1304
Db       1110 ENLSEEEKNILYEKKASAWYHVTYHPEWVKKS-----------LELQ---DPDESSHAAM 1155

Qy       1305 RASTAFKLYYHKSPKFVWQMAGRQLAYIK 1333
Db       1156 LS-------------FAW-IAADYLARIK 1170
```

RESULT 4
Q9LKP0
ID Q9LKP0 PRELIMINARY; PRT; 1196 AA.
AC Q9LKP0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE RNA-dependent RNA polymerase.
GN SDE1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20306669; PubMed=10850496;
RA Dalmay T., Hamilton A., Rudd S., Angell S., Baulcombe D.C.;
RT "An RNA-dependent RNA polymerase gene in Arabidopsis is required for
RT posttranscriptional gene silencing mediated by a transgene but not by
RT a virus.";
RL Cell 101:543-553(2000).
DR EMBL; AF268093; AAF74208.1; -.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR InterPro; IPR007855; RdRP.
DR Pfam; PF05183; RdRP; 1.
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1196 AA; 136747 MW; C1B1961CDC11F708 CRC64;

Query Match 4.4%; Score 326; DB 10; Length 1196;
Best Local Similarity 22.5%; Pred. No. 4.4e-13;
Matches 200; Conservative 119; Mismatches 297; Indels 272; Gaps 42;

```
Qy        535 FFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPL 594
Db        466 FVAPIVKDLTSSSFSQKTYVFKRVKSI------LTDG-FKLCGRKY-SFLAFSAN----- 512

Qy        595 REFQLRAEDPKPIIKERVHFFAETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWL 654
Db        513 ---QLR--------DRSAWFFAEDG--------KTR----------------VSDIKTWM 537

Qy        655 LQL-DNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGR 713
Db        538 GKFKDKNV----AKCAARMGLCFSSTYATVDVMPHEV---DTEVPDIERNGYVFSDGIGT 590

Qy        714 MSRSVAKRIRDVLGLGDV---PSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWEC 770
Db        591 ITPDLAGEVMEKLKL-DVHYSPCAYQIRYAGFKGV-------------VARWPS------ 630

Qy        771 DFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLIND------- 823
Db        631 -------------------KSDGIRLAL--------RD--SMKKIFSKHTILEICSWTRF 661

Qy        824 ----LQRQFSEQKHALNRPVEFRQWVYESYSSRATRV---SHGRVPFLAGLPDSQEETLN 876
Db        662 QPGFLNRQIITLLSVLGVPVEIFWDMQESMLYKLDRILVDTDVAFEVLTASCAEQGNTAA 721

Qy        877 FLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVH 936
Db        722 IMLSAGFKPKTEPHLRGMLSSVRIAQLWGLREKSRIFVTSGRWLMGCLDEAGILEHGQCF 781

Qy        937 VGFS------------SKFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHS 984
Db        782 IQVSKPSIENCFSKHGSRFKETKKDLEVVKGY-VAIAKNPCLHPGDVRILEAVDVPQLHH 840

Qy        985 LKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRY--LKK 1042
Db        841 MYDCLIFPQKGDRPHTNEASGSDLDGDLYFVAWDQKLI----------PPNRKSYPAMHY 890

Qy       1043 DKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTN----YKERLCYINNS 1098
Db        891 DAAEEKSL------GRAVNHQDIIDFFARN----LANEQLGTICNAHVVHADRSEY---G 937

Qy       1099 VSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPD-PMYKSDSWLGRG 1157
Db        938 AMDEECLLLAELAATAVDFPKTGKIVSMP--FHLKPKLYPDFMGKEDYQTYKSNKILGR- 994

Qy       1158 EPTHIIDYLKFSIARP---AIDKELEA--------------------FHNAMKAA---K 1190
Db        995 ------------LYRPVKEVYDEDAEASSEESTDPSAIPYDAVLEIPGFEDLIPEAWGHK 1042

Qy       1191 DTEDG-------------------AHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRI 1231
Db       1043 CSYDGQLIGLLGQYKVQKEEEIVTGHIWS---MPKYTSKKQGELKER--------LKHSY 1091

Qy       1232 GEVEKEYGRLVKNKEMRDSKDPYPVRVNQVYEK----WCAIT--PEAMDKSGANYDSKVI 1285
Db       1092 NSLKKEF-RKVFEETIPDHENLSEEEKNILYEKKASAWYHVTYHPEWVKKS--------- 1141

Qy       1286 RLLELSFLADREMNTWALLRASTAFKLYYHKSPKFVWQMAGRQLAYIK 1333
Db       1142 --LELQ---DPDESSHAAMLS-------------FAW-IAADYLARIK 1170
```

RESULT 5
Q8LHH9
ID Q8LHH9 PRELIMINARY; PRT; 1218 AA.
AC Q8LHH9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative RNA-directed RNA polymerase.
GN B1074C08.21.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT clone:B1074C08.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP004357; BAC00725.1; -.
DR Gramene; Q8LHH9; -.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR InterPro; IPR007855; RdRP.
DR Pfam; PF05183; RdRP; 1.
KW RNA-directed RNA polymerase.
SQ SEQUENCE 1218 AA; 136533 MW; 984FD38DC5474482 CRC64;

Query Match 4.4%; Score 325; DB 10; Length 1218;
Best Local Similarity 21.6%; Pred. No. 5.3e-13;
Matches 199; Conservative 149; Mismatches 318; Indels 254; Gaps 43;

```
Qy        497 VLDYNPDNSPTAPLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSK 556
Db        482 VLNNNVLNSFTAPI----VKDLM------------SNFFQ-----------------Q 506
```

```
Qy    557 QPGAVEEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFA 616
Db    507 KTTVYKRVRMLLTEGFH-MCGRKY-SFLAFSSN--------QLR--------DKSAWFFA 548

Qy    617 ETGITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGL 676
Db    549 E---------------------DRKT--TVEAIRKWMGRFTSKNVAKHA---ARMGQCF 581

Qy    677 SKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGD-VPSAV 735
Db    582 SSTYATVTMRPDEVDESFDDVVH---NEYIFSDGIGKITPDLALEVAERLQLTDNPPSAY 638

Qy    736 QGRFGSAKGMWVIDV-DDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGL 794
Db    639 QIRFAGFKG--VIAVWQGHGDGTRLFLRPSMRKFESNHL-----VLEVVS-WTKFQPGFL 690

Qy    795 NLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRAT 854
Db    691 NRQIIILLSS---------LNVPDSIFWQMQETMLSN---LNNILSDRDVAFEVLTTSCA 738

Qy    855 RVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRV 914
Db    739 ---------------DDGNTAALMLSAGFEPRTEPHLKAMLLAIRSAQLQDLLEKARIFV 783

Qy    915 GRSAYIYMIADFWGVLEE-----------NEVHVGFSSKFRDEEESFTLLSDCDVLVAR 962
Db    784 PKGRWLMGCLDELGVLEQGQCFIRATVPSLNSYFVKHGSRFSSTDKNTEVILG-TVVIAK 842

Qy    963 SPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIV 1022
Db    843 NPCLHPGDVRILEAVDVPELHHLVDCLVFPQKGERPHANEASGSDLDGDLYFVTWDEKLI 902

Qy   1023 DGFVNAEMPLE---PDLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQP 1079
Db    903 PPGKKSWNPMDYSPPEA-----------KQL---------PRQVSQHDIIDFFLKNMISE 942

Qy   1080 NFLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGG 1139
Db    943 NLGRICNAHVVHADLSEYGAMDEKCIHLAELAATAVDFPKTG-----------KLAIMPP 991

Qy   1140 ALS---LPDPM-------YKSDSWLGR----------GEPTHIIDYLKFSIARP-----A 1174
Db    992 HLKPKVYPDFMGKEDGQSYKSEKILGRLYRSIQEASNG------DVVSQEVCTPNDLPYD 1045

Qy   1175 IDKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEV 1234
Db   1046 IDLEVPGASDFLASAWQCKCS---YDAQLSALLSQY-----RVRTEAELVT------GHI 1091

Qy   1235 EKEYGRLVKN--KEMRDSKDPYPVRVNQVYEKWCAITPEAMDKS-GANYDSKVIRLLELS 1291
Db   1092 T----FLVKNSSKKQGDIKD----RLKTAY--------SALRKEFKSTFESIASDQCE-- 1133

Qy   1292 FLADREMNTWALLRASTAFKLYYHKSPKFVWQMAGRQLAYIKAQMTSRPGEGAPALMTAF 1351
Db   1134 -IGDDEKNLLYEMKASAWYQVTYH--PKWVEKSRG---------ILGPDGEEIPASLS-- 1179

Qy   1352 MYAGLMPDKKFTKQYVARLE 1371
Db   1180 -FAWIPVD------YLARIK 1192
```

```
RESULT 6
Q9ZRX3
ID   Q9ZRX3      PRELIMINARY;      PRT;    775 AA.
AC   Q9ZRX3;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   RNA-directed RNA polymerase (Fragment).
GN   RDRP.
OS   Petunia hybrida (Petunia).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Solanales; Solanaceae; Petunia.
OX   NCBI_TaxID=4102;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leaf;
RA   Schiebel W., Pelissier T., Riedel L., Thalmeir S., Schiebel R.,
RA   Kempe D., Lottspeich F., Saenger H.L., Wassenegger M.;
RT   "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from
RT   Tomato.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ011979; CAA09896.1; -.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   InterPro; IPR007855; RdRP.
DR   Pfam; PF05183; RdRP; 1.
KW   RNA-directed RNA polymerase.
FT   NON_TER       1      1
FT   NON_TER     775    775
SQ   SEQUENCE   775 AA;  88778 MW;  DE43FE50C6A45B93 CRC64;

  Query Match             4.4%;  Score 322.5;  DB 10;  Length 775;
  Best Local Similarity   21.2%;  Pred. No. 3.8e-13;
  Matches  174;  Conservative  119;  Mismatches  311;  Indels  217;  Gaps   32;

Qy    562 EEVIQWLTMGQHSLVGRQWRAFFAKD---AGYRKPLREFQLRAEDPKPIIKERVHFFAET 618
Db     80 EDIDNFLRV---SFVDEEWEKIFSMDLLPRANRNGIR----------TDIYERILSTLQN 126

Qy    619 GITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLD---W-------LLQLDNNTWQPHL-- 666
Db    127 GLVIGAKKFEFLA-------------FSSSQLRDNSVWMFASRPGLTAYDIRTWMGDFQQ 173

Qy    667 -----KLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKR 721
Db    174 IRNVAKYAARLGQSFGSSRETLSVGRHEIKYIPD--VEVHGTNYVFSDGIGKISADFAHR 231

Qy    722 IRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLE 781
Db    232 VAKKCGLQHTPSAFQIRYGGYKG--VVAVDPYSPMK-LSLRKSMLKYE-----SNNTKLD 283

Qy    782 VRSVASELKSAGLNLQLLPVLED-RARDKV---KMRQAIG--DRLINDLQRQFSEQKHAL 835
Db    284 VLG-WSKYQPCYLNRQLVTLMSTLGVKDEVFKRKQNEAVAQLDAILHDSLK----AQEAL 338

Qy    836 NRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIA 895
Db    339 D---------------------------LMSPGENTNILKEMLNCGYLPDAEPFLSMML 370

Qy    896 WDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFSS----KFRDEEESF- 950
Db    371 QTFRASKLLDLRTRTRIFIPKGRSMMGCLDESRTLEYGQVFVQFTGAGGRQFYEESHHFN 430

Qy    951 -------TLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPLAKKL 1003
Db    431 DSGSANYNFILKGNVVVAKNPCLHPGDIRVLKAVDVPALHHMVDCVVFPQKGKRPHPNEC 490

Qy   1004 SGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTT------FKQLMASHGTG 1057
Db    491 SGSDLDGDIYFVCWDPDLIPPRQVQPMDYTPAPSIQLDHDVTIEEVEEYFTNYMVNDGLG 550

Qy   1058 SAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQ 1117
Db    551 VIANAHVVF--------ADREPNM-----------------AMSDPCIELAQLFSIAVDF 585

Qy   1118 SKQGIVFNEASWAQLRRELLGGALSLPD-PMYKSDSWLGRGEPTHIIDYLKFSIARPAID 1176
Db    586 PKTGVPAEIPS--QLRPKEYPDFMEKPDKTTYHSEKVIGK-------------LFREVKD 630

Qy   1177 KELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEV-- 1234
Db    631 KTTQA----------------------SSITTFTRDVARRSYDTDLEVDGFEDYIDEAFD 668

Qy   1235 -EKEYGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDKSGA----------NYDSK 1283
```

```
Db        669 YKSEY-----DKKLGNLMDYYGIKT------------EAETLSGGIMKASKTFDRRRDAE 711

Qy       1284 VIRLLELSFLADREMNTW---------ALLRASTAFKLYYH 1315

Db        712 AIGVAVRSL--RKEARTWFKRRSDIDDLLAKASAWYHVAYH 750

RESULT 7
Q7XM31
ID   Q7XM31      PRELIMINARY;      PRT;  1136 AA.
AC   Q7XM31;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   OSJNBb0048E02.13 protein.
GN   OSJNBB0048E02.13.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA   Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA   Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA   Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA   Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA   Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA   Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA   Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA   Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL606653; CAE04828.1; -.
SQ   SEQUENCE   1136 AA;  126924 MW;  308B542CCB45AB09 CRC64;

  Query Match             4.3%;  Score 320.5;  DB 10;  Length 1136;
  Best Local Similarity   22.9%;  Pred. No. 9.7e-13;
  Matches  160;  Conservative  115;  Mismatches  248;  Indels  175;  Gaps   32;

Qy        698 LSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLG--DVPSAVQGRFGSAKGMWVIDVDDTGD 755

Db        539 ITTDGSKYIFSDGIGKISLRFAKRVAHHVGLDPTNLPSAFQIRYGGYKG--VIAIDPMSS 596

Qy        756 EDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRA-RDK--VKM 812

Db        597 ID-LSLRPSMKKFESE-----SRMLNITS-WSKSQPCYVNREIISLLSTLGIRDEIFVAM 649

Qy        813 RQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQE 872

Db        650 QQ--------DEMRETEEM--LTNKEV-------------------ALSVLGKLGGSET 679

Qy        873 ETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEE 932

Db        680 KTAVKMLLQGYEPSSEPYLSMILKAHQENRLTDIRTRCKIHVPKGRVLIGCLDETGVLEY 739

Qy        933 NEVHVGF---SSKFRDEEESFTLLSD-------CDVLVARSPAHFPSDIQRVRAVFKPEL 982

Db        740 GQVYIRITKNSKEQKDSNQSYFYNDDGKTATVVGKVAITKNPCLHPGDIRVLEAIYDPDL 799

Qy        983 HSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKK 1042

Db        800 VGMVDCLVFPQRGERPHPNECSGGDLDGDLYFITWDDKLIPEKVDTPM------------ 847

Qy       1043 DKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNS---V 1099

Db        848 DYTATRPRIMDH--------VVTLEEIQKHFVDYMINDSLGAIST--AHLIHADRSPLKA 897

Qy       1100 SNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPD-------PMYKSDS 1152

Db        898 RSPECLQLATLHSMAVDFAKTG------APAEMPRTL--RPREYPDFMERWEKPMYISNG 949

Qy       1153 WLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWDPDL--------- 1203

Db        950 VLGK-------------LYRSAMG-HMEKSGDSGALSSSSAQPSPTYDPDLEVPGSDEFL 995

Qy       1204 ---ASYYTFFKEISDKSRSSALLFTTLKNRI-GEVEKEY--GRLVKNKEMRDSKDPYPVR 1257

Db        996 QAAEEYYELYEE----------KLTTLMNYYRAELEDEILTGN-IRNKML------YLKR 1038

Qy       1258 VNQVYEKWCAITPEAMDKSGANYDSKVIRLLELSFLADREMNTWAL---------LRAST 1308

Db       1039 DNKRYF-------EMKDRIVAAVDA-----------LHREARGWLLSSRKEEDASRMASA 1080

Qy       1309 AFKLYYHKSPK-------FVWQMAGRQLAYIKAQMTSR 1339

Db       1081 WYRVTYHPDRRRGKRFWSFPW-IACDNLLAIKASSQLR 1117

RESULT 8
Q872R5
ID   Q872R5      PRELIMINARY;      PRT;  1351 AA.
AC   Q872R5;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Related to RNA-directed RNA polymerase.
GN   B13B3.100.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX284762; CAD70515.1; -.
DR   GO; GO:0016329; F:apoptosis regulator activity; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0006915; P:apoptosis; IEA.
DR   InterPro; IPR000712; Bcl2_BH.
DR   InterPro; IPR007855; RdRP.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF05183; RdRP; 1.
DR   Pfam; PF02205; WH2; 1.
DR   PROSITE; PS01259; BH3; 1.
KW   RNA-directed RNA polymerase.
SQ   SEQUENCE   1351 AA;  151975 MW;  0C784E99CDC66AE6 CRC64;

  Query Match             4.2%;  Score 312;  DB 3;  Length 1351;
  Best Local Similarity   22.5%;  Pred. No. 4.9e-12;
  Matches  246;  Conservative  126;  Mismatches  369;  Indels  350;  Gaps   54;

Qy        299 FEAPPSQPREK----------RP-------VDATVFEAGH---------LIESPSKGRTT 332

Db         92 YEAEEKERRKKAWEENTSNNARPARTPTLALDADVLDCGHWSYIDGNLTFITEWSSSLRV 151

Qy        333 KSHIDNQPLSSSSQGETSFSTYY----ESFPSSGGEGAIPEPSRSNGLARSEESARSQVQ 388

Db        152 SAKFAKHDLVITSLAESQVSIPYRYIHELVWSNGGHVAV-------------------TL 192

Qy        389 VHAPV-VAARLRNIWPKFPKWLH-EAPLAVAWEVTR--LFMHCKVDLEDESL-------- 436

Db        193 AHAPTFLSPPIPLLRPGQPKRMRLEACDSNHAKVSKFCLVYHFKV--SDKHLKQNQGSYR 250

Qy        437 GLKYDPSWSTARDVTDIWKTLYRLDAFRGKPFPEKPPNDVFVTAMTGNFESKGSAVVLSA 496

Db        251 GSDFRTAINGLNEQEMFWVTHYTF------AIQTTSPQDAYAVAVDR---------LRAA 295

Qy        497 VLDYNPDNS------------------PTAPLYLVKLKPLMFEQGCRLTRRFGP----- 532
```

```
Db        296 LLEYEGRDTLPYSLLFNLQALVYMSYLHPTTVLKLAKRLADMFETARRSGQRQDPISVDA 355

Qy        533 --DRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMG--QH-SLVGRQWRAFFAKD 587
Db        356 FKDLFKTIDWPSPSLSS------------AEELAQFEVEGIIEHLKKTEKRMREGYTLR 402

Qy        588 AG---------YRKPLREFQLRAEDPKPIIKERV---------HF----FAETGITFRP 624
Db        403 LNEEIPPGLTKIYRALVTPTRIELHGPELEAKNRILRKFPEHQDHFLRVQFAEED---GQ 459

Qy        625 DVFKTRSVVPAEEPVEQRTEFK-----------VSQMLD-----------WL------- 654
Db        460 DLFFNSAV--SMDAIYQR--FKDVLTNGISVGGRVYRFLGFSHSSLRAHSLWLAAPFIYD 515

Qy        655 --LQLDNNTWQ--------PHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLL---SP 700
Db        516 GKLQLASNIIEDLGDFRNIMSPARRAARIGQAFSETPYSVSLYDHGI-----DVIRQRDV 570

Qy        701 SGTGEVMNDGVGRMSRSVAKRI-RDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWI 759
Db        571 KRNERVFSDGVGIISQGALEVIHREIPESKGYPNCLQVRWAGAKGMLALDARLTGRQICI 630

Qy        760 ETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDR 819
Db        631 R--DSMEK----FRSRDEEHLEICDMASKPIPLMLNRQMIKILED-------------- 669

Qy        820 LINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFL- 878
Db        670 --------------MRAPA---QWFLELQEKELQR--------LRAITDNVQNVATFLK 703

Qy        879 ------------------MNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAY 919
Db        704 LQCVGDSVHLSQFLKDLDKMN--IDYRRDQFLRGIVEAVVLRELRLLKHKARIPVPYGVT 761

Qy        920 IYMIADFWGVLEENEVHVGFSS---KFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRA 976
Db        762 LFGVMDETGLLREGEVYVTFETVDGRFKDPPTA------GPVVVTRSPALHPGDIQIAHN 815

Qy        977 VFKP---ELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIV---DGFVNAEM 1030
Db        816 AIPPAGHPLRELKNCIVFSQNGERDLPSQLSGGDLDGDTFNVIWDQSIVATLRTFAAADY 875

Qy       1031 P-LEP-DLSRYLK-KDKTTF--KQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMC 1085
Db        876 PRVEPLKLNREVESKDMADFFVEFMKADHLGVIAVRHMILAD----------------- 917

Qy       1086 TNYKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPD 1145
Db        918 ----ER----NEGTLDADCLKLAALHSKAVDFSKSGI---HVDITELPR---------P 956

Qy       1146 PMYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWDPDLAS 1205
Db        957 PMYRPD-FLVNGPDIKIHD-------KSTIDMEEQYLR------QDDDDG------DDTP 996

Qy       1206 YYTFFKEISDK 1216
Db        997 RYKYYK--SDK 1005
```

RESULT 9
Q95ZG6
ID Q95ZG6 PRELIMINARY; PRT; 1964 AA.
AC Q95ZG6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative RNA dependent RNA polymerase (Fragment).
GN RRPB.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Martens H., Novotny J., Oberstrass J., Steck T.L., Postlethwait P.,
RA Nellen W.;
RT "RNAi in Dictyostelium: developmental regulation and the role of RdRPs
RT and dsRNase.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ314910; CAC41975.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001201; PAP_25A_core.
DR InterPro; IPR007855; RdRP.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF05183; RdRP; 1.
DR SMART; SM00490; HELICc; 1.
KW ATP-binding; Helicase; Hydrolase.
FT NON_TER 1 1
SQ SEQUENCE 1964 AA; 227413 MW; B83CA9098675E645 CRC64;

Query Match 4.0%; Score 298; DB 5; Length 1964;
Best Local Similarity 20.0%; Pred. No. 8.4e-11;
Matches 173; Conservative 136; Mismatches 291; Indels 266; Gaps 39;

```
Qy        509 PLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWL 568
Db        772 PSRTIFIQPTL-QRSCRAVRKFGSSNFIMVKI-----VNESLEPL-------------- 810

Qy        569 TMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFRPDVFK 628
Db        811 ----------------QGVNNKHPLHENRI-----KPIL--------EGGILVGGKIYS 840

Qy        629 TRSVVPAEEPVEQRTEFKVSQMLDWLL--QLDNNT---WQ--PHL----KLFSRIQLGLS 677
Db        841 YAG----------NSNSQLREYSSWFVSNQIGTHTVKIWSGIEHVDNVRKFFRCIGLMFS 890

Qy        678 KTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQG 737
Db        891 TTIPTVTLPQNRI-YRIQDI---TRNTHVFTEGCGEIGPELAKHLNENYNFRPSTCAYQV 946

Qy        738 RFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQ 797
Db        947 RIGGNKGMLVVN-NQAPDPSGIYIRPSMVKFNPIVCGDEHRTLEICSV-STTSRCKLNRQ 1004

Qy        798 LLPVLED-RARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRV 856
Db       1005 VISLLSTLGTQDNV----------------FFALQDHYLN---QVAQIVNDTNASKQAIV 1045

Qy        857 SHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGR 916
Db       1046 EF--------FPDITEGEL----------YQDPYIRRILISLYKLKMERIQQKCHIEIKD 1087

Qy        917 SAYIYMIADFWGVLEENEVHVGFSSKFRDEEE--SFTLLSDCDVLVARSPAHFPSDIQRV 974
Db       1088 SRMLLGVCDPTNSLPPNTVFVQLEEEDEDDDGRKYEKVIEGLVMVIKNPCTHPGDVRYL 1147

Qy        975 RAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEP 1034
Db       1148 KAVDNLRLRHLRNVLVFSTKGDVPNFKEISGSDLDGDRYFFCYDKSLI------------ 1195

Qy       1035 DLSRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLG---MCTNYKER 1091
Db       1196 ----------------------GNRSKSETAY---------------LGDETVSNNDKKA 1218

Qy       1092 LCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGA-----LSLPDP 1146
Db       1219 ------NVFNDP-FALSSMYSTNVERQELGKMYN----SHLAISDLFGANHEFSIQISKE 1267

Qy       1147 MYKSDSWLGRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDTEDGAHFWDPDLAS- 1205
Db       1268 CFKE------------IDYPKTGI-HGTIPKEANVW-------LKTKGYPHYMQRENSTR 1307
```

```
Qy       1206 -YY-------------------TFFKEIS-DKS----------RSSALLFTTLKNRIGE 1233
Db       1308 VYYQSKTIMGKMYDQIDQLVYIGDFLPNISLDKSNLVDDYEIYLNSAKILYSQYKLQVHS 1367

Qy       1234 VEKEYGR--------------LVKNKEMRD----SKDPYPVRVNQVYEKWCAITPEAMDK 1275
Db       1368 LLRHYSAESEESIMIGFLDQGFISDKVSKDIKGEMKNDY-IKIQQTFEN------EFLKE 1420

Qy       1276 SGANYDSKVIRLLELSFLADREMNTW 1301
Db       1421 FGEQHKENCLLIHRVNI--EKKVSAW 1444

RESULT 10
Q9ZR58
ID   Q9ZR58      PRELIMINARY;      PRT;  1114 AA.
AC   Q9ZR58;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   RNA-directed RNA polymerase (EC 2.7.7.48).
GN   RDRP.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rutgers;
RX   MEDLINE=99055198; PubMed=9836747;
RA   Schiebel W., Pelissier T., Riedel L., Thalmeir S., Schiebel R.,
RA   Kempe D., Lottspeich F., Sanger H.L., Wassenegger M.;
RT   "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from
RT   tomato.";
RL   Plant Cell 10:2087-2102(1998).
DR   EMBL; Y10403; CAA71421.1; -.
DR   PIR; T30819; T30819.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   InterPro; IPR007855; RdRP.
DR   Pfam; PF05183; RdRP; 1.
KW   Nucleotidyltransferase; RNA-directed RNA polymerase; Transferase.
SQ   SEQUENCE   1114 AA;  126807 MW;  B6D30ACD41DB37FF CRC64;

  Query Match             4.0%;  Score 297;  DB 10;  Length 1114;
  Best Local Similarity   20.4%;  Pred. No. 3.9e-11;
  Matches  175;  Conservative  133;  Mismatches  323;  Indels  228;  Gaps   32;

Qy        562 EEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGIT 621
Db        386 EDIDNFLRV---SFVDEEWEKLYSTDL--------------LPK----------ASTGSG 418

Qy        622 FRPDVFK------TRSVVPAEEPVEQRTEFKVSQMLDWLLQLDNNTW----QPHLKL--- 668
Db        419 VRTNIYERILSTLRKGFVIGDKKFE-FLAFSSSQL------RDNSVWMFASRPGLTANDI 471

Qy        669 ------FSRI--------QLGLSKTYAIMTLEPHQIRHHKTDLL---SPSGTGEVMNDGV 711
Db        472 RAWMGDFSQIKNVAKYAARLGQSFGSSRETL---SVLRHEIEVIPDVKVHGTSYVFSDGI 528

Qy        712 GRMSRSVAKRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECD 771
Db        529 GKISGDFAHRVASKCGLQYTPSAFQIRYGGYKGVVGVDPDSSMK---LSLRKSMSKYESD 585

Qy        772 FVDKHQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQ 831
Db        586 NI-----KLDVLG-WSKYQPCYLNRQLITLLS---------TLGVKDEVLEQKQKEAVDQ 630

Qy        832 KHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYL 891
Db        631 LDAILHDSLKAQEALELMS-----------------PGENTNILKAMLNCGYKPDAEPFL 673

Qy        892 QDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIADFWGVLEENEVHVGFS----SKFRDEE 947
Db        674 SMMLQTFRASKLLDLRTRSRIFIPNGRTMMGCLDESRTLEYGQVFVQFTGAGHGEFSDDL 733

Qy        948 ESF--------TLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFSTKGDVPL 999
Db        734 HPFNNSRSTNSNFILKGNVVVAKNPCLHPGDIRVLKAVNVRALHHMVDCVVFPQKGKRPH 793

Qy       1000 AKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQLMASHGTGSA 1059
Db        794 PNECSGSDLDGDIYFVCWDQDMIPPRQVQPMEYPPAPSIQLDHD---------------- 837

Qy       1060 AKEQTTYDMIQKSFHFALQPNFLGMCTN----YKERLCYINNSVSNKPAIILSSLVGNLV 1115
Db        838 ----VTIEEVEEYFTNYIVNDSLGIIANAHVVFADR---EPDMAMSDPCKKLAELFSIAV 890

Qy       1116 DQSKQGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLKFSIARPAI 1175
Db        891 DFPKTGVPAEIPS--QLRPK------EYPDFMDKPD------KTSYISERVIGKLFRKVK 936

Qy       1176 DKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFTTLKNRIGEV- 1234
Db        937 DKA----------------------PQASSIATFTRDVARRSYDADMEVDGFEDYIDEAF 974

Qy       1235 --EKEYGRLVKNKEMRDSKDPYPVRVNQVYEKWCAITPEAMDKSGANYD--------SKV 1284
Db        975 DYKTEY-----DNKLGNLMDYYGIKTE------AEILSGGIMKASKTFDRRKDAEAISVA 1023

Qy       1285 IRLLELSFLADREMNTW---------ALLRASTAFKLYYHKS--------------PKFV 1321
Db       1024 VRALR------KEARAWFKRRNDIDDMLPKASAWYHVTYHPTYWGCYNQGLKRAHFISFP 1077

Qy       1322 WQMAGRQLAYIKAQMTSRP 1340
Db       1078 WCVYDQLIQIKKDKARNRP 1096

RESULT 11
Q19285
ID   Q19285      PRELIMINARY;      PRT;  1780 AA.
AC   Q19285; Q22232;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   F10B5.7 protein.
GN   F10B5.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Burton J.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Sims M.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z66500; CAA91312.1; -.
DR   EMBL; Z48334; CAA91312.1; JOINED.
DR   EMBL; Z48334; CAA88315.1; -.
DR   EMBL; Z66500; CAA88315.1; JOINED.
DR   PIR; T20695; T20695.
DR   WormPep; F10B5.7; CE02178.
DR   InterPro; IPR007855; RdRP.
DR   Pfam; PF05183; RdRP; 1.
SQ   SEQUENCE   1780 AA;  203145 MW;  16444F29CEF902C2 CRC64;

  Query Match             4.0%;  Score 296.5;  DB 5;  Length 1780;
  Best Local Similarity   20.4%;  Pred. No. 9e-11;
  Matches  196;  Conservative  129;  Mismatches  348;  Indels  289;  Gaps   37;
```

```
Qy    468 FPEKPPNDVFVTAMTGNFES---------KGSAVVLSAVLDYNPDNSPT------APLYL 512
Db    631 FDERRQIDV-VTAFTTMYQSRKIQYERLLKGESLQDVGLAKPLPKNCVSVAKVIVTPSRI 689

Qy    513 VKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWLTMGQ 572
Db    690 LLMAPEVMMVN-RVVRRFGPD--------------------------------------- 709

Qy    573 HSLVGRQWRAFFAKDAGYRKPLREFQLRAED-PKPIIKERVHFFAETGITFRPDVFK--- 628
Db    710 YAL-----RCVFRDDNLGRLAIRDFSINNIDHMSNIVTEGIYLTLKNGIQVADRVYSFLG 764

Qy    629 ---------------------TRSVVPAEEPV-----EQRTEFKVSQMLDWLLQLDNNTW 662
Db    765 WSNSQMRDQGCYLYAPRVNALTGEVTGTVEDIRVWMGDFRDAISVPKMMSRMGQCFTQA- 823

Qy    663 QPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMN-----DGVGRMSRS 717
Db    824 QP------TVYSSVKNIHIVENIQVRLERHHW--IVEPDIEGGVENKYCFSDGCGRISIK 875

Qy    718 VAKRIRDVLGLGDVPSAVQGRFGSAKGMWVID--VDDTGDEDWIETYPSQRKWECDFVDK 775
Db    876 LATHISKILQLKEVPACFQVRFKGFKGILVIDPTIDDIINMPKVIFRKSQQKFGEGGGEL 935

Qy    776 HQRTLEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQ--------------------- 814
Db    936 QDEYLEVVKYAMP-SPVCLNRPFITIL-----DQVSEKQSASSHRRITNRVHYYLERELC 989

Qy    815 AIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEET 874
Db    990 SLSNMLINENQ----AAEELVNRTNLAIDW--NAASKRA--------------------- 1022

Qy    875 LNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNI----RVGRSAYIYMIADFWGVL 930
Db   1023 -------GFELSVDPLIRDMLFSIYRYNIIHHISKAKIFLPPSLGRS--MYGVVDETGLL 1073

Qy    931 EENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVII 990
Db   1074 QYGQVFIQYSPSIR-QTSNRPILKTGKVLITKNPCHVPGDVRVFDAVWQPALAHLVDVVV 1132

Qy    991 FSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDKTTFKQL 1050
Db   1133 FPQHGPRPHPDEMAGSDLDGDEYSIIWDQEMLLDY-NEEAMVFP---------------- 1175

Qy   1051 MASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYINNS---VSNKPAIIL 1107
Db   1176 --SSSAAEEDKEPTTDDMVEFFLRY-LQQDSIGRMSH--AHLAYADLHGLFHENCHAIAL 1230

Qy   1108 SSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDPMYKSDSWLGRGEPTHIIDYLK 1167
Db   1231 KCAVA--VDFPKSGVPAEPLS-------------SFEQCEMTPDYMMSGGKPMYYSTRLN 1275

Qy   1168 FSIARPA--IDKELEAFHNAMKAAKDTEDGAHFWDPDLASYYTFFKEISDKSRSSALLFT 1225
Db   1276 GQLHRKARKVEEVLEEF---------------------------------ETRGSVF--- 1299

Qy   1226 TLKNRIGEVEKEYGRL-------------VKNKEMRDSKDPYPVRVNQVYEKWCAITPEA 1272
Db   1300 ---------EREYDKLICPEDVDVFFGNEIKLVQTLTLRDEYVDRMQQLLDEY-GIEDEA 1349

Qy   1273 MDKSGANYDSKVIRLLELSFLADREMNTWALLRASTAFKLYYHK-----SPKFVWQMAGR 1327
Db   1350 SVVSG--------HAASIKRLAGMERDDYSFYHTDKVVELRYEKLYAVFRAKFFEEFGGE 1401

Qy   1328 QL 1329
Db   1402 EI 1403
```

```
RESULT 12
Q9LQV2
ID   Q9LQV2      PRELIMINARY;      PRT;  1107 AA.
AC   Q9LQV2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   F10B6.19.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chao Q., Shinn P., Dunn P., Buehler E., Kahn S., Kim C., Walker M.,
RA   Williams S., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
RA   Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S.,
RA   Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,
RA   Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RT   "Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome
RT   I.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ecker J.R.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Ecker J.R.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA   Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
RA   Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA   Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA   Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA   Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA   Theologis A., Ecker J.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006917; AAF79241.1; -.
DR   GO; GO:0005853; C:eukaryotic translation elongation factor 1 . . . .; IEA.
DR   GO; GO:0003746; F:translation elongation factor activity; IEA.
DR   GO; GO:0006414; P:translational elongation; IEA.
DR   InterPro; IPR001326; EF1_BD.
DR   InterPro; IPR007855; RdRP.
DR   Pfam; PF05183; RdRP; 1.
SQ   SEQUENCE   1107 AA;  126214 MW;  F8FF97691FF82187 CRC64;

  Query Match             4.0%;  Score 295;  DB 10;  Length 1107;
  Best Local Similarity   21.4%;  Pred. No. 5.2e-11;
  Matches 131;  Conservative  105;  Mismatches  240;  Indels  136;  Gaps   21;

Qy    638 PVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDL 697
Db    470 PIDRIT---AAHIRAWMGDFD------HIRNVAKYAARLGQSFS-SSRETLNVRSDEIEV 519

Qy    698 LS-----PSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDV-PSAVQGRFGSAKGMWVIDVD 751
Db    520 IPDVEIISLGTRYVFSDGIGKISAEFARKVARKCGLTEFSPSAFQIRYGGYKGVVAVDPN 579

Qy    752 DTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLL------PVLEDR 805
Db    580 SS-----------------------KKLSLRKSMSKFESENTKLDVLAWSKYQPCYMNR 615

Qy    806 ARDKVKMRQAIGDRLINDLQRQFSEQKHA-LNRPVEFRQWVYESYSSRATRVSHGRVPFL 864
Db    616 QLITLLSTLGVTDSVFEKKQREVVDRLDAILTHPLE----------------AHEALGLM 659

Qy    865 AGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIA 924
Db    660 A--PGENTNILKALILCGYKPDAEPFLSMMLQNFRASKLLELRTKTRIFISGGRSMMGCL 717

Qy    925 DFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHS 984
Db    718 DETRTLEYGQVVVQYSDPMRPGRR---FIITGPVVVAKNPCLHPGDVRVLQAVNVPALNH 774
```

```
Qy        985 LKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDK 1044
Db        775 MVDCVVFPQKGLRPHPNECSGSDLDGDIYFVCWDQELVPPRTSEPMDYTPEPTQILDHD- 833
Qy       1045 TTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNY------KERLCYINNS 1098
Db        834 -------------------VTIEEVEEYFANYIVNDSLGIIANAHTAFADKEPL-----K 869
Qy       1099 VSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRREL----LGGALSLPD-PMYKSDSW 1153
Db        870 AFSDPCIELAKKFSTAVDFPKTGVA------AVIPQHLYVKEYPDFMEKPDKPTYESKNV 923
Qy       1154 LG--------RGEPTHIIDYLKFSIARPAIDKEL----------EAFHNAMKAAKDTEDG 1195
Db        924 IGKLFREVKERAPPLISIKSFTLDVASKSYDKDMEVDGFEEYVDEAFY--QKANYDFKLG 981
Qy       1196 AHFWDPDLASYY 1207
Db        982 ------NLMDYY 987
```

RESULT 13
Q8TGV4
ID   Q8TGV4   PRELIMINARY;   PRT; 1484 AA.
AC   Q8TGV4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative RNA-dependent RNA polymerase RDP-1.
GN   RDP-1.
OS   Diaporthe perjuncta.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Diaporthales; Valsaceae; Diaporthe.
OX   NCBI_TaxID=186170;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CMW3407;
RA   Preisig O., Wingfield M.J.;
RT   "Putative fungal RNA-dependent RNA polymerase.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF468822; AAL78034.1; -.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   InterPro; IPR007855; RdRP.
DR   Pfam; PF05183; RdRP; 1.
KW   RNA-directed RNA polymerase.
SQ   SEQUENCE   1484 AA; 167213 MW; 57E0AFB8D12547BD CRC64;

Query Match 4.0%; Score 294.5; DB 3; Length 1484;
Best Local Similarity 21.6%; Pred. No. 9.2e-11;
Matches 268; Conservative 153; Mismatches 444; Indels 373; Gaps 67;

```
Qy        136 NTPGRTFGRTLSGPSGLSRPTST----NTKRKDEPANVTFADPPKRSLTRSATGPPIHGA 191
Db        106 NASARTF-----GPRGHPNPRQNRFMPNRRQQRSP----------QSLQQRSTPAPLTGA 150
Qy        192 AIP---------LKFPDPVNTGSKRPSLESENLN----QCTKRAKGKLSDNVAAAAAPPV 238
Db        151 VLSAAQSLTIDLFGLPTDIKIREIWESMTSNHLNVNRVEIYENFDGKRTGKARVIIEPP- 209
Qy        239 PIASALDKVPTRRHANTRDPTATGHRRADQVDSFDTSQGTSYGSSVFSACRHNQSTTQSS 298
Db        210 PI---------------RWPTWSG---PDNLITID--------------RADSSSLQVS 236
Qy        299 FEAPPSQPREK----RPVDATVFEAGHLIESPSK----GRTTKSHIDNQP---------- 340
Db        237 LETRPFYPSTRTPCNRPIPAQFCQKKLL---PTKLVFGMMASESAMLNMPQFSYEDVVLL 293
Qy        341 -LSSSSQGETSFSTYY-----ESFPS---------------SGGEGA----IPEPSRSNG 375
Db        294 TKFSKRQFELQFTVNLGRYGCKAFKSQIDFSHPTKIIWGRRENGAWALAMLLPSPPKLYG 353
Qy        376 ---LARSEESARSQV------QVHAPVVAARLRNI--------WP--KFPKWL-----HE 411
Db        354 RLPVSHDDEQTLWQEWDLWARQTNISSEMHKLRDLPAKLEKQTWKSIEFGRWTTYYLDME 413
Qy        412 APLAVAWEVTRLFM---HCKVDLEDESLGLKYDPSWSTARDVTDIWKTL----------- 457
Db        414 SSLS-EWESVRTYLADYNVKI-LQVDHFDL-VNPGLSNA------WKLLDNSAEMTLELL 464
Qy        458 ----YRLDAFRGKPFPEKPPNDVFVTAMTGNFESKG-SAVVLSAVLDYNPDNSPTAPLY 511
Db        465 GSTQSYHLD------FPVRYQLEVCIS--HGLLSEYGVNAQFLEKLISFETDRA------ 510
Qy        512 LVKLKPLMFEQGCRLTRR-FGPDRFFE---ILIPSPTSTSPSVPPVVSK---QPGAV--- 561
Db        511 -----RMMLEGVAEANKQFFEPMEMFDDPRILHYWPNARIPPYATLVRRAVITPTTIYFK 565
Qy        562 ---EEVIQWLTMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAET 618
Db        566 TPCVELTNRILRKYSDLNDRFLRVQFTDEVTFGKIFSSQDSNKDDN---LYTRVHRVMQN 622
Qy        619 GITFRPDVFKTRSVVPAEEPVEQRTEFKVSQMLD----WLLQLDNNT------WQ---PH 665
Db        623 GVIIGDRHYKFLA-------------FSNSQFRENGAFFFCETDHVTCASIRNWMGDFRH 669
Qy        666 L----KLFSRIQLGLSKTYAIMTLEPHQIRHHKTDLLSPSGTGEVMN--DGVGRMSRSVA 719
Db        670 IRSVGKFAARMGQCFTTTRQVNGISIPKIR-QIGDIQRRSG-DNIWNFTDGVGKISVFFA 727
Qy        720 KRIRDVLGLGDVPSAVQGRFGSAKGMWVIDVDDTGDEDWIETYPSQRKWECDFVDKHQRT 779
Db        728 RMIASERDLPETPSCFQMRMGGCKGVVVVWPDIPANE--VHIRPSQEKFKAVY-----NG 780
Qy        780 LEVRSVASELKSAGLNLQLLPVLEDRARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPV 839
Db        781 LEIIK-TSTFAHATLNKQVIPVLIALGVD-----DAVFVRMLDDELKEYDE---AL---- 827
Qy        840 EFRQWVYESYSSRATRVSHGRVPFLAGLPDSQEETLNFL-MNSGFDPKKQKYLQDIAWDL 898
Db        828 --------ADSMKAGELLRSQV-------DENQTTLTMAEMVDTFMDSEEPFL----WTL 868
Qy        899 QKR-KC---DTLKSKLNIRVGRSAYIYMIADFWGVLEEN--------------------E 934
Db        869 LRLWKCWVLKRLKQKFAISVKNSAMIFGVVDETGVLRGHSQDTEGKGFNNIESLPQIFLQ 928
Qy        935 VHV-GFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHSLKDVIIFST 993
Db        929 VPIEGSDGKSTTNYEVITGI----CVVGRNPALHPGDVRVVEAVDVPELRHLKNVVVFPK 984
Qy        994 KGDVPLAKKLSGGDYDGDMAWVCWDPEIV---------DGFVNAEMP--------LEPDL 1036
Db        985 TGDRDVPSMCSGGDMDGDDYFVYWDERLIPSEWDHPPLDHDADSSTPGLDSPADVTNGDV 1044
Qy       1037 SRYLKKDKTTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNYKERLCYIN 1096
Db       1045 TRF-------FAQYMKNDSLGRIA-----------TAHFAQADQLAG------------- 1073
Qy       1097 NSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGALSLPDPMY--KSDSWL 1154
Db       1074 -GVKHPKCIELAKLHSKAVDYIKSG------KPAVMKRHL--QPRNWPHWMERDKKSSYH 1124
Qy       1155 GRGEPTHIIDYLKFSIARPAIDKELEAFHNAMKAAKDT 1192
Db       1125 SRGALGQQIYDRIK-----------IEEFHAAYEMSFDT 1151
```

RESULT 14
Q95ZG7
ID   Q95ZG7   PRELIMINARY;   PRT; 1883 AA.
AC   Q95ZG7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative RNA dependent RNA polymerase (Fragment).
GN   RRPA.

```
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Martens H., Novotny J., Oberstrass J., Steck T.L., Postlethwait P.,
RA   Nellen W.;
RT   "RNAi in Dictyostelium: developmental regulation and the role of RdRPs
RT   and dsRNase.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ314909; CAC41974.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR001201; PAP_25A_core.
DR   InterPro; IPR007855; RdRP.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF05183; RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
KW   ATP-binding; Helicase; Hydrolase.
FT   NON_TER    1883   1883
SQ   SEQUENCE   1883 AA;  217454 MW;  1B1D9B9738B15094 CRC64;

  Query Match             4.0%;  Score 293.5;  DB 5;  Length 1883;
  Best Local Similarity   20.5%;  Pred. No. 1.6e-10;
  Matches  173;  Conservative  145;  Mismatches  305;  Indels  221;  Gaps   41;

Qy        509 PLYLVKLKPLMFEQGCRLTRRFGPDRFFEILIPSPTSTSPSVPPVVSKQPGAVEEVIQWL 568
Db        947 PSRTIFIQPTL-QRSCRAVRKFGSSNFIVVKI-----VNESLEPL--------------- 985

Qy        569 TMGQHSLVGRQWRAFFAKDAGYRKPLREFQLRAEDPKPIIKERVHFFAETGITFRPDVFK 628
Db        986 ----------------QGVNNKHPLHENRI-----KPIL--------EGGILVGGEIYS 1015

Qy        629 TRSVVPAEEPVEQRTEFKVSQMLDWLL--QLDNNT---WQ--PHL----KLFSRIQLGLS 677
Db       1016 YAG---------NSNSQLREYSSWFVSNQIGTHTVKIWSGIEHVDNVRKFFRCIGLMFS 1065

Qy        678 KTYAIMTLEPHQIRHHKTDLLSPSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDVPSAVQG 737
Db       1066 TTIPTVTLPQNRI-YRIQDI--TRNTHE-FTEGCGEIGPELAKHLNENYNFRPSTCAYQV 1121

Qy        738 RFGSAKGMWVIDVDDTGDEDWIETYPSQRKWE-CDFVDKHQRTLEVRSVASELKSAGLNL 796
Db       1122 RIGGNKGVLVVN-NQAPDPSGIYIRPSMVKFNPIDCGDEH-RTLEICSV-STTSRCKLNR 1178

Qy        797 QLLPVLED-RARDKVKMRQAIGDRLINDLQRQFSEQKHALNRPVEFRQWVYESYSSRATR 855
Db       1179 QVISLLSTLGTQDNV---------------FFALQDHYLN---QVAQIVNDTNASKQAI 1219

Qy        856 VSHGRVPFLAGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVG 915
Db       1220 VEF--------FPDITEGEL----------YQDPYIRRILISLYKLKMERIQQKCHIEIK 1261

Qy        916 RSAYIYMIADFWGVLEENEVHVGFSSKFRDEEE---SFTLLSDCDVLVARSPAHFPSDIQ 972
Db       1262 DSRMLLGVCDPTNSLPPNTVFVQLEEEDEDDDDGRKYEKVIEGLVMVIKNPCTHPGDVR 1321

Qy        973 RVRAVFKPELHSLKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPL 1032
Db       1322 YLKAVDNIRLRHLRNVLVFSTKGDVPNFKEISGSDLDGDRYFFCYDKSLIGNRSESE--- 1378

Qy       1033 EPDLSRYL--------KKDKTTFKQ---LMASHGTGSAAKEQTTYDMIQKSFHFALQPNF 1081
Db       1379 ----TAYLVVETVSNNDKKANVFNDPFALSSMYSTNAERQE------LGKLYH-----SH 1423

Qy       1082 LGMCTNYKERLCYINNSVSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRRELLGGAL 1141
Db       1424 LAISDQY---------GANHKYSIQISKECFKEIDYPKTGIHGTIPKEVNIRLQTVG--- 1471

Qy       1142 SLPDPM---------YKSDSWLGR-----GEPTHIIDYL-KFSIARPAIDKELEAFHNAM 1186
Db       1472 -YPHYMQRENSTRVYYQSKKIMGKMYDQIDQLVYIGDFLPNISLDKSNLVDGYEIYLNSA 1530

Qy       1187 KAAKDTEDGAHFWDPDLASYY-----TFFKEISDKSRSSALLFTTLKNRIGEVEKEYGRL 1241
Db       1531 KI-------------LYSQYKLQVHSLLRHYSAESEESIM--------IGFLDQGF--- 1565

Qy       1242 VKNKEMRD----SKDPYPVRVNQVYEKWCAITPEAMDKSGANYDSKVIRLLELSFLADRE 1297
Db       1566 ISDKVSKDIKGEMKNDY-IKIQQTFEN------EFLKEFGEQHKENCLLIHRVNI--EKK 1616

Qy       1298 MNTW 1301
Db       1617 VSAW 1620
```

```
RESULT 15
Q8H1K9
ID   Q8H1K9     PRELIMINARY;      PRT;  1107 AA.
AC   Q8H1K9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RNA-dependent RNA polymerase 1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yu D., Fan B., MacFarlane S.A., Chen Z.;
RT   "Analysis of the involvement of an inducible Arabidopsis RNA-dependent
RT   RNA polymerase in antiviral defense.";
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY148431; AAN64409.1; -.
DR   GO; GO:0005853; C:eukaryotic translation elongation factor 1 . . .; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0003746; F:translation elongation factor activity; IEA.
DR   GO; GO:0006414; P:translational elongation; IEA.
DR   InterPro; IPR001326; EF1_BD.
DR   InterPro; IPR007855; RdRP.
DR   Pfam; PF05183; RdRP; 1.
KW   RNA-directed RNA polymerase.
SQ   SEQUENCE   1107 AA;  126204 MW;  C52CFCA1E5880CB7 CRC64;

  Query Match             3.9%;  Score 287;  DB 10;  Length 1107;
  Best Local Similarity   21.2%;  Pred. No. 1.9e-10;
  Matches  130;  Conservative  105;  Mismatches  241;  Indels  136;  Gaps   21;

Qy        638 PVEQRTEFKVSQMLDWLLQLDNNTWQPHLKLFSRIQLGLSKTYAIMTLEPHQIRHHKTDL 697
Db        470 PIDRIT---AAHIRAWMGDFD------HIRNVAKYAARLGQSFS-SSRETLNVRSDEIEV 519

Qy        698 LS-----PSGTGEVMNDGVGRMSRSVAKRIRDVLGLGDV-PSAVQGRFGSAKGMWVIDVD 751
Db        520 IPDVEIISLGTRYVFSDGIGKISAEFARKVARKCGLTEFSPSAFQIRYGGYKGVVAVDPN 579

Qy        752 DTGDEDWIETYPSQRKWECDFVDKHQRTLEVRSVASELKSAGLNLQLL------PVLEDR 805
Db        580 SS------------------------KKLSLRKSMSKFESENTKLDVLAWSKYQPCYMNR 615
```

```
Qy    806 ARDKVKMRQAIGDRLINDLQRQFSEQKHA-LNRPVEFRQWVYESYSSRATRVSHGRVPFL 864

Db    616 QLITLLSTLGVTDSVFEKKQREVVDRLDAILTHPLE----------------AHEALGLM 659

Qy    865 AGLPDSQEETLNFLMNSGFDPKKQKYLQDIAWDLQKRKCDTLKSKLNIRVGRSAYIYMIA 924

Db    660 A--PGENTNILKALILCGYKPDAEPFLSMMLQNFRASKLLELRTKTRIFISGGRSMMGCL 717

Qy    925 DFWGVLEENEVHVGFSSKFRDEEESFTLLSDCDVLVARSPAHFPSDIQRVRAVFKPELHS 984

Db    718 DETRTLEYGQVVVQYSDPMRPGRR---FIITGPVVVAKNPCLHSGDVRVLQAVNVPALNH 774

Qy    985 LKDVIIFSTKGDVPLAKKLSGGDYDGDMAWVCWDPEIVDGFVNAEMPLEPDLSRYLKKDK 1044

Db    775 MVDCVVFPQKGLRPHPNECSGSDLDGDIYFVCWDQELVPPRTSEPMDYTPEPTQILDHD- 833

Qy   1045 TTFKQLMASHGTGSAAKEQTTYDMIQKSFHFALQPNFLGMCTNY------KERLCYINNS 1098

Db    834 -------------------VTIEEVEEYFANYIVNDSLGIIANAHTAFADKEPL-----K 869

Qy   1099 VSNKPAIILSSLVGNLVDQSKQGIVFNEASWAQLRREL----LGGALSLPD-PMYKSDSW 1153

Db    870 AFSDPCIELAKKFSTAVDFPKTGVA------AVIPQHLYVKEYPDFMEKPDKPTYESKNV 923

Qy   1154 LG--------RGEPTHIIDYLKFSIARPAIDKEL----------EAFHNAMKAAKDTEDG 1195

Db    924 IGKLFREVKERAPPLISIKSFTLDVASKSYDKDMEVDGFEEYVDEAFY--QKANYDFKLG 981

Qy   1196 AHFWDPDLASYY 1207

Db    982 ------NLMDYY 987
```

Search completed: March 29, 2004, 08:49:19
Job time : 63 secs

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         March 31, 2004, 06:26:16 ; Search time 5456.8 Seconds
                                           (without alignments)
                                           7672.409 Million cell updates/sec

Title:          US-09-913-878A-2
Perfect score:  7397
Sequence:       1 MNPITPRKRNSPVEEIINRL..........YEVLGDDDFDGIGFTGNGDY 1402

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters:      55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09913878/runat_29032004_082008_8379/app_query.fasta_1.2254
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09913878_@CGN_1_1_4565_@runat_29032004_082008_8379 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      EST:*
                1:  em_estba:*
                2:  em_esthum:*
                3:  em_estin:*
                4:  em_estmu:*
                5:  em_estov:*
                6:  em_estpl:*
                7:  em_estro:*
                8:  em_htc:*
                9:  gb_est1:*
                10:  gb_est2:*
                11:  gb_htc:*
                12:  gb_est3:*
                13:  gb_est4:*
                14:  gb_est5:*
                15:  em_estfun:*
                16:  em_estom:*
                17:  em_gss_hum:*
                18:  em_gss_inv:*
                19:  em_gss_pln:*
                20:  em_gss_vrt:*
                21:  em_gss_fun:*
                22:  em_gss_mam:*
                23:  em_gss_mus:*
                24:  em_gss_pro:*
                25:  em_gss_rod:*
                26:  em_gss_phg:*
                27:  em_gss_vrl:*
                28:  gb_gss1:*
                29:  gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	266	3.6	2199	11	AY103827	AY103827 Zea mays
	2	204	2.8	883	14	CA765940	CA765940 AF53-Rpf_
	3	203.5	2.8	972	14	CF885460	CF885460 tric082xd
	4	197.5	2.7	787	14	CB907399	CB907399 tric082xd
	5	191.5	2.6	577	10	BF096404	BF096404 EST360431
	6	188.5	2.5	3758	11	BC015736	BC015736 Homo sapi
	7	186.5	2.5	737	14	CF883391	CF883391 tric088xh
	8	186.5	2.5	776	14	CB909520	CB909520 tric088xh
c	9	184.5	2.5	954	29	CG445102	CG445102 OG8AH57TV
	10	183.5	2.5	641	14	CA257764	CA257764 SCCCFL110
	11	175.5	2.4	582	12	BI246074	BI246074 IP1_68_D1
	12	173	2.3	622	13	BU965311	BU965311 sat08g07.
	13	172	2.3	761	10	BF275078	BF275078 GA__Eb002
	14	171	2.3	934	29	CG212111	CG212111 OG0FJ51TV
	15	170	2.3	3267	11	BC036553	BC036553 Homo sapi
	16	169.5	2.3	648	14	CA223511	CA223511 SCJFFL1C0
	17	169.5	2.3	741	14	CF441078	CF441078 EST677423
	18	169	2.3	3990	11	BC039311	BC039311 Homo sapi
	19	168	2.3	493	13	BU067573	BU067573 1626_E08_
	20	168	2.3	541	13	BQ827639	BQ827639 gd58f03.y
	21	168	2.3	677	14	CD928156	CD928156 GR45.104C
	22	167.5	2.3	916	29	CG223171	CG223171 OGWAL14TV
	23	166	2.2	722	14	CF134248	CF134248 WHE4369_F
	24	166	2.2	812	29	CG140649	CG140649 PUIHY65TB
	25	164	2.2	570	10	BE433646	BE433646 EST400175
	26	163.5	2.2	585	14	CF244986	CF244986 3530_1_8_
	27	163.5	2.2	4117	11	BC012130	BC012130 Homo sapi
	28	163	2.2	640	10	BF460205	BF460205 069G02 Ma
c	29	162.5	2.2	1891	29	CG756328	CG756328 P051-3-E0
c	30	161	2.2	786	29	CG809201	CG809201 FSAAE20TF
	31	161	2.2	3312	11	AK050985	AK050985 Mus muscu
	32	161	2.2	6166	29	AY417941	AY417941 Mus muscu
	33	160.5	2.2	3423	11	AK084676	AK084676 Mus muscu
c	34	159.5	2.2	741	13	BQ997497	BQ997497 QGG16J23.
	35	158	2.1	5220	29	AY413592	AY413592 Homo sapi
	36	157.5	2.1	621	13	BU997627	BU997627 HI08I16r
	37	156	2.1	1487	12	BG923047	BG923047 602821859
	38	154.5	2.1	629	12	BJ268864	BJ268864 BJ268864
c	39	154	2.1	781	29	CG938103	CG938103 MBENF75TR
	40	154	2.1	1379	10	BF133062	BF133062 601645474
c	41	154	2.1	1704	29	CG756741	CG756741 P051-4-F0
c	42	153.5	2.1	872	29	CG266238	CG266238 OGXAZ41TH
	43	153.5	2.1	885	28	BH163571	BH163571 ENTSB82TF
c	44	152.5	2.1	893	29	CC699849	CC699849 OGKAJ24TH
	45	152	2.1	2072	11	BC048977	BC048977 Homo sapi

ALIGNMENTS

```
RESULT 1
AY103827
LOCUS       AY103827                2199 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO135426 mRNA sequence.
ACCESSION   AY103827
VERSION     AY103827.1  GI:21206905
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2199)
```

```
AUTHORS     Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE       Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
JOURNAL     Unpublished (2002)
REFERENCE   2  (bases 1 to 2199)
AUTHORS     Coe,E.H.
TITLE       Direct Submission
JOURNAL     Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..2199
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:637927"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN

Alignment Scores:
Pred. No.:               2.63e-09      Length:        2199
Score:                   266.00        Matches:       148
Percent Similarity:      37.96%        Conservative:  109
Best Local Similarity:   21.86%        Mismatches:    278
Query Match:             3.60%         Indels:        142
DB:                      11            Gaps:          24

US-09-913-878A-2 (1-1402) x AY103827 (1-2199)

Qy     674 LeuGlyLeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHis 693
           ||||||   |||   :::   :::         |||||||:::   |||:::
Db      66 CTTGGGCAATCTTTCAGTTCCTCAACAGAAACTTTAAAAGTACACAAATCTGAGGTGGAA 125

Qy     694 LysThrAspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArg 713
           :::          ::::::   :::||||||      :::   :::|||||||||||||:::
Db     126 CGAATTCCTGATATTACA---AATGGCACAAAGTACATATTCTCTGATGGAGTTGGAAAG 182

Qy     714 MetSerArgSerValAla------------------------LysArgIleArgAsp 724
           :::|||   :::    |||                        |||||
Db     183 ATCTCAGCTAATTTTGCAGTGGAGGTGGCTATGAAGTGCAAATTGAAACGCTTT------ 236

Qy     725 ValLeuGlyLeuGlyAspValProSerAlaValGlnGlyArgPheGlySerAlaLysGly 744
                                   ||||||   |||   |||:::|||       ||||||
Db     237 -----------------GCTCCTTCTGTTTTCAGATAAGGTATGGCGGTTACAAAGGT 278

Qy     745 MetTrpValIleAspValAspAspThrGlyAspGluAspTrpIleGluThrTyrProSer 764
           :::       :::         ||||||            :::            |||
Db     279 GTTGTCGCTGTA---------GATACAAGATCAAATCATAAGCTTTCTTTGAGAAAAAGC 329

Qy     765 GlnArgLysTrpGluCysAspPheValAspLysHisGlnArgThrLeuGluValArgSer 784
                  |||::::::   :::   :::               |||||||:::|||   :::
Db     330 ATGTCAAAGTTCCAGTCAGAAAATATC---------------ACTCTTGATGTCCTTGCA 374

Qy     785 ValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAsp 804
                 |||:::   :::         ||||||   ||||||:::   :::|||
Db     375 TAC---AGCAAGTACCAACCATGCTTCCTGAATCGGCAGTTGATTACTCTTCTCTCA--- 428

Qy     805 ArgAlaArgAspLysValLysMetArgGlnAlaIleGlyAspArgLeuIleAsnAspLeu 824
                                         :::   |||   :::

Db     429 -----------------------ACACTTGGGGTTAGCGATAATGTCTTTGAGCTAAAG 464

Qy     825 GlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGlnTrp 844
           |||::::::          |||         |||||||||   |||
Db     465 CAGAAGGAAGCCTTAAGGCAG---------TTGAACAGAATGGTAACT----------- 503

Qy     845 ValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPheLeu 864
                          :::|||   |||   :::         :::|||
Db     504 ---------------GAACCACAGGCTGCTCGTGAAGCAGTTGAACTTATGCCCATG--- 545

Qy     865 AlaGlyLeuProAspSerGlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAsp 884
                                       :::      |||::::::   |||:::
Db     546 ------------GGAGAAGTAACCAATGTAGTTAAAGAATTGTTGTCATGTGGCTACCAG 593

Qy     885 ProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAsp 904
           |||      :::   ||||||         :::         :::   |||
Db     594 CCTGATCATGAGCCATATCTTTCCATGCTGCTACAAACTTTTAGAGCATCCAAGCTTCTA 653

Qy     905 ThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAla 924
              ||||||:::|||   |||   :::   :::         :::
Db     654 GAGTTGAAAACAAAGTCAAGGATATTCATCACACAGGGGCGAGCAATGATGGGTTGCCTG 713

Qy     925 AspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPheSerSerLysPheArg 944
           |||         |||:::   :::|||   |||         |||
Db     714 GATGAAACCTGCACACTTAAGTACGGCCAGGTATTCGTCCAAGCTTCTTACAGTGCAGAT 773

Qy     945 AspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerPro 964
           |||      :::         ::::::         |||:::||||||:::::|||
Db     774 GACCATCGCAAG---------GTCGTTGTAACTGGAAAAGTAGTTGTCGCCAAAAATCCT 824

Qy     965 AlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSer 984
                 |||   ||||||:::   ::::::||||||         |||   ||||||
Db     825 TGTCTCCACCCTGGTGACATACGGGTTCTCCAGGCTGTTGATGTTCCTGCTCTGCACCAC 884

Qy     985 LeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSer 1004
           |||   |||   ::::::|||      :::|||         |||      :::   |||
Db     885 TTGTTTGACTGTGTTGTCTTTCCACAGCAGGGACCAAGGCCGCACCCTAATGAGTGTTCA 944

Qy    1005 GlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGly 1024
           |||   |||   |||||||||:::   :::|||   |||||||||   ::::::
Db     945 GGGAGTGATCTTGATGGGGACATATATTTTGTTTCTTGGGATCCACATCTTATTCCAAGT 1004

Qy    1025 PheValAsnAlaGluMetProLeuGluProAspLeuSerArgTyrLeuLysLysAspLys 1044
              :::      |||      |||   :::   |||      |||
Db    1005 CGTTTGGTGGATCCTATGGACTATACTCCAGCTTCAGCAGAAACATTAGACCATGAT--- 1061

Qy    1045 ThrThrPheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThr 1064
Db    1062 ---------------------------------------------------------GTC 1064

Qy    1065 ThrTyrAspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPheLeuGlyMet 1084
           |||   :::   ||||||:::   |||            :::      ||||||:::
Db    1065 ACTATTGAGGAGATACAGGAGTACTTCACAAACTACATAGTTAATGAGAGTCTTGGGATT 1124

Qy    1085 CysThrAsnTyr------------------LysGluArgLeuCysTyrIleAsnAsnSer 1098
              |||                        :::|||||||:::
Db    1125 ATCGCCAATGCGCATGTGGTCTTTACAGATCAGGAACGTATG---------------AAA 1169

Qy    1099 ValSerAsnLysProAlaIleIleLeuSerSerLeuValGlyAsnLeuValAspGlnSer 1118
              :::   |||   :::   |||:::   |||            ||||||
Db    1170 GCTGAGAGTCCACCGTGCGTTCAACTGGCCAAGCTCTTCTCTATAGCTGTCGATTTCCCA 1229

Qy    1119 LysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeuArgArgGluLeuLeuGly 1138
           |||   |||:::                        |||   :::   ||||||
Db    1230 AAGACTGGAGTGCCG------------------GCTCTGATTCCACATGAGCTA------ 1265

Qy    1139 GlyAlaLeuSerLeuProAspProMet---------------------TyrLysSerAsp 1151
                 ||||||   |||                              |||:::|||
Db    1266 CATGTCAAGGAGTATCCTGACTTCATGGAGAAACTCGACAAAGTCACCTATGAATCAAAG 1325

Qy    1152 SerTrpLeuGlyArg---------------GlyGluProThrHisIleIleAspTyrLeu 1166
              :::|||:::                   :::   |||      |||   :::
Db    1326 GGTGTGATCGGGAAGCTCTATAGGGAAATAAAGAAGCACACACCACACATAAAGCACTTC 1385
```

```
Qy       1167 LysPheSerIleAlaArgProAlaIleAspLysGluLeu------GluAlaPheHisAsn 1184
Db       1386 ACGAGGGAAGTGGCAAGGCGGTCTTATGACACCGATTTGATTGTTGATGGCTATGAAGAT 1445

Qy       1185 AlaMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeuAla 1204
Db       1446 TACATTACTGAGGCTATAGAGTTCAAGGAAGAGTACGATTTCAGGCTGGGTAATCTTATG 1505

Qy       1205 SerTyrTyrThrPhePheLysGlu---------IleSerAspLysSerArgSerSerAla 1221
Db       1506 GACCACTATGGCATAAAAAGTGAAGCTGAGATAATAAGTGGATGTATTCTAAAGATGGCA 1565

Qy       1222 LeuLeuPheThr------------------ThrLeuLysAsnArgIleGlyGluValGlu 1235
Db       1566 AAGAATTTCACCAAGAGTAGTGATGCTGATGCAATTAGAATGGCGGTGAGATCTTTGAGG 1625

Qy       1236 LysGlu-----------TyrGlyArgLeuValLysAsnLysGluMetArgAspSerLys 1251
Db       1626 AAAGAAGCTAGGTCGTGGTTCAATGAGATGAGCACAGGAGAGGATGGCCAAGATGCCATG 1685

Qy       1252 AspPro---------------TyrProValArgValAsnGlnValTyr------------ 1262
Db       1686 GAGGCCAAGGCCTCTGCTTGGTACCATGTTACTTATCATCAGCAGTACTGGGGCAGCTAC 1745

Qy       1263 --------------------------------GluLysTrpCysAlaIleThrProGlu 1271
Db       1746 AATGAAGGGTATGATCGGCCGCATCTTATTAGCTTCCCATGGTGCGTATATGACAAGCTT 1805

Qy       1272 AlaMetAspLysSerGlyAlaAsnTyrAspSerLysValIleArgLeuLeu 1288
Db       1806 GTGGCCATCAAGCAGGGGAGGAATCTCCTCACGCAGATGGATCGAAACTTG 1856
```

```
RESULT 2
CA765940
LOCUS       CA765940               883 bp    mRNA    linear   EST 08-JAN-2003
DEFINITION  AF53-Rpf_09_H23_T7_092.ab1 IRRI Drought Stress Panicle Library
            Oryza sativa (indica cultivar-group) cDNA clone C0003263 5' similar
            to unknown, mRNA sequence.
ACCESSION   CA765940
VERSION     CA765940.2  GI:27547887
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 883)
  AUTHORS   Bennett,J., Arumugam,K., Lafitte,R., Wen,J., Rudd,S. and
            Bruskiewich,R.M.
  TITLE     IRRI Drought Stress Panicle cDNA Library
  JOURNAL   Unpublished (2002)
COMMENT     On Dec 2, 2002 this sequence version replaced gi:25995195.
            Contact: Richard Bruskiewich
            Biometrics and Bioinformatics Unit
            International Rice Research Institute
            DAPO 7777, Metro Manila, Philippines
            Tel: +63-2-845-0563
            Fax: +63-2-845-0606
            Email: r.bruskiewich@cgiar.org
            International Rice Information System (IRIS;
            http://www.iris.irri.org): D0203262
            Assignment of putative function to the sequence by S. Rudd of the
            Munich Information Center for Protein Sequences
            (http://mips.gsf.de)
            Plate: 09  row: H  column: 23.
FEATURES             Location/Qualifiers
     source          1..883
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="IR64"
                     /db_xref="taxon:39946"
                     /clone="C0003263"
                     /tissue_type="Panicles"
                     /dev_stage="Flowering"
                     /clone_lib="IRRI Drought Stress Panicle Library"
                     /note="Vector: pBluescript II SK+; Water stress was
                     applied by not watering for 4 consecutive days. Panicles
                     were collected from control (well watered) and stressed
                     plants at 2 days before heading, at heading, 50% flowering
                     and 4 days after 50% flowering."
ORIGIN

Alignment Scores:
Pred. No.:              3.67e-05      Length:         883
Score:                  204.00        Matches:        56
Percent Similarity:     46.26%        Conservative:   43
Best Local Similarity:  26.17%        Mismatches:     99
Query Match:            2.76%         Indels:         16
DB:                     14            Gaps:           3

US-09-913-878A-2 (1-1402) x CA765940 (1-883)
```

```
Qy        824 LeuGlnArgGlnPheSerGluGlnLysHisAlaLeuAsnArgProValGluPheArgGln 843
Db         45 TTAAATCGACAGATTATAATATTGCTATCCTCACTGAATGTCCCAGATTCTATCTTTTGG 104

Qy        844 TrpValTyrGluSerTyrSerSerArgAlaThrArgValSerHisGlyArgValProPhe 863
Db        105 CAAATGCAAGAGACCATGCTTTCTAACCTCAACAATATTCTATCAGACAGAGATGTTGCT 164

Qy        864 LeuAlaGlyLeuProAspSer---------GlnGluGluThrLeuAsnPheLeuMetAsn 880
Db        165 TTTGAGGTTTTAACAACTTCATGTGCTGATGATGGAAACACTGCAGCATTGATGCTCAGT 224

Qy        881 SerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAspLeuGlnLys 900
Db        225 GCTGGCTTTGAACCTAGAACTGAACCACACTTGAAAGCAATGCTCTTGGCTATAAGGTCT 284

Qy        901 ArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTyrIle 920
Db        285 GCACAATTGCAGGATCTTTTGGAAAAAGCAAGGATATTTGTGCCAAAGGGAAGGTGGTTG 344

Qy        921 TyrMetIleAlaAspPheTrpGlyValLeuGluGlu------------------------ 932
Db        345 ATGGGCTGTCTTGATGAGCTTGGGGTTCTTGAGCAAGGGCAGTGCTTTATTCGGGCAACA 404

Qy        933 ------------AsnGluValHisValGlyPheSerSerLysPheArgAspGluGluGlu 948
Db        405 GTTCCATCATTGAATAGTTATTTTGTTAAGCATGGTCAAGATTTTCATCAACAGATAAA 464

Qy        949 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 968
Db        465 AACACAGAGGTCATTTTGGGA---ACTGTGGTAATAGCAAAGAATCCCTGTCTTCATCCA 521

Qy        969 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 988
Db        522 GGGGATGTCCGCATCCTTGAAGCAGTTGATGTGCCCGAACTGCATCATCTGGTTGATTGT 581

Qy        989 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyr 1008
Db        582 TTGGTGTTCCCCCAGAAAGGTGAGAGGCCACATGCTAACGAGGCATCTGGGAGCGATCTT 641

Qy       1009 AspGlyAspMetAlaTrpValCysTrpAspProGluIleVal 1022
Db        642 GATGGGGATCTCTACTTTGTGACATGGGATGAGAAACTTATA 683
```

```
RESULT 3
CF885460
LOCUS       CF885460               972 bp    mRNA    linear   EST 31-OCT-2003
DEFINITION  tric082xd19.b11 T.reesei mycelial culture, Version 6 October 2003
            Hypocrea jecorina cDNA clone tric082xd19, mRNA sequence.
ACCESSION   CF885460
VERSION     CF885460.1  GI:38140142
KEYWORDS    EST.
SOURCE      Hypocrea jecorina (anamorph: Trichoderma reesei)
```

```
  ORGANISM  Hypocrea jecorina
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE   1  (bases 1 to 972)
  AUTHORS   Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,
            Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and
            Dean,R.A.
  TITLE     Analysis of the protein processing and secretion pathways in a
            Trichoderma reesei EST dataset
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Ralph A. Dean
            Fungal Genomics Laboratory
            North Carolina State University
            Campus Box 7251, Raleigh, NC 27695, USA
            Tel: 919-513-0020
            Fax: 919-513-0024
            Email: ralph_dean@ncsu.edu
            Seq primer: LT-F1 primer.
FEATURES             Location/Qualifiers
     source          1..972
                     /organism="Hypocrea jecorina"
                     /mol_type="mRNA"
                     /strain="QM6a"
                     /db_xref="taxon:51453"
                     /clone="tric082xd19"
                     /dev_stage="mycelia"
                     /clone_lib="T.reesei mycelial culture, Version 6 October
                     2003"
                     /note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial
                     culture grown from 24 hrs to 6 days with varying Carbon
                     and Nitrogen sources and concentrations."
ORIGIN

Alignment Scores:
Pred. No.:              4.64e-05     Length:         972
Score:                  203.50       Matches:        51
Percent Similarity:     48.25%       Conservative:   18
Best Local Similarity:  35.66%       Mismatches:     69
Query Match:            2.75%        Indels:         5
DB:                     14           Gaps:           1

US-09-913-878A-2 (1-1402) x CF885460 (1-972)

Qy    1248 ArgAspSerLysAspProTyrProValArgValAsnGlnValTyrGluLysTrpCysAla 1267
           ||||||   |||         :::      :::         ||||||:::||||||
Db       6 CGGGATTGAAAACTAAAGTTTGCCAACAAGGCCGAGGCTGTCTACCAAAAATGGCTGGAC 65

Qy    1268 IleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysValIleArgLeu 1287
           |||   |||                             :::   :::   |||
Db      66 ATACAACCGCCGCCCGAGCTGCTGATGAGCGGCGAGTTTCCCGCGCTAGTCCATGACGCA 125

Qy    1288 LeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrpAlaLeuLeuArgAlaSer 1307
              :::   |||   |||::::::                  |||   ||||||:::||||||
Db     126 AACAAGGCTTCGGCGTTGAGTAAC---------------TGGCAGCTGCTGAAAGCTTCG 170

Qy    1308 ThrAlaPheLysLeuTyrTyrHisLysSerProLysPheValTrpGlnMetAlaGlyArg 1327
           |||   ||||||   |||:::|||   :::|||||||||||||||   |||||||||:::
Db     171 ACCACTTTCAAGAAGTATCACCACAGCGCGCCCAAGTTTGTCTGGAACATGGCCGGAAAG 230

Qy    1328 GlnLeuAlaTyrIleLysAlaGlnMetThrSerArgProGlyGluGlyAlaProAlaLeu 1347
           |||||||||||||||||||||         :::      ||||||   :::      |||
Db     231 CAGCTTGCCTACATCAAGGCGATGGCGCGAAACGACGCTGGCGAGACGTCAAAGGTGCTT 290

Qy    1348 MetThrAlaPheMetTyrAlaGlyLeuMetProAspLysLysPheThrLysGlnTyrVal 1367
           :::            |||:::      |||   ||||||||||||
Db     291 GTGATTCCCGAGATGTGGGGGGTCCTGCGGCCGGACAAGAAGCTCATCACCTCGCTGGCG 350

Qy    1368 AlaArgLeuGluGlyAspGlySerGluTyrProAspProGluValTyrGluValLeuGly 1387
              :::   |||                              :::   ||||||
Db     351 ATACAGAGGGAGGCTGCGCGGGATTCGGAGAGTGCACTGGCTTTGGAGGAGGTGTTTGAG 410

Qy    1388 AspAspAsp 1390
              ||||||
```

```
Db       411 TTCGACGAT 419

RESULT 4
CB907399
LOCUS       CB907399                 787 bp    mRNA    linear   EST 02-JUL-2003
DEFINITION  tric082xd19 T.reesei mycelial culture, Version 3 april Hypocrea
            jecorina cDNA clone tric082xd19, mRNA sequence.
ACCESSION   CB907399
VERSION     CB907399.1  GI:30122057
KEYWORDS    EST.
SOURCE      Hypocrea jecorina (anamorph: Trichoderma reesei)
  ORGANISM  Hypocrea jecorina
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE   1  (bases 1 to 787)
  AUTHORS   Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
            Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
            Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
            Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
  TITLE     Transcriptional regulation of biomass-degrading enzymes in the
            filamentous fungus Trichoderma reesei
  JOURNAL   J. Biol. Chem. 278 (34), 31988-31997 (2003)
  MEDLINE   22803314
   PUBMED   12788920
COMMENT     Contact: Pamela K. Foreman
            Genencor Intl.
            925 Page Mill Road, Palo Alto, CA 94304, USA
            Tel: (650) 846-7635
            Fax: (650) 621-7817
            Email: Pforeman@genencor.com
            Seq primer: LT-F1 primer.
FEATURES             Location/Qualifiers
     source          1..787
                     /organism="Hypocrea jecorina"
                     /mol_type="mRNA"
                     /strain="QM6a"
                     /db_xref="taxon:51453"
                     /clone="tric082xd19"
                     /dev_stage="mycelia"
                     /clone_lib="T.reesei mycelial culture, Version 3 april"
                     /note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial
                     culture grown from 24 hrs to 6 days with varying Carbon
                     and Nitrogen sources and concentrations."
ORIGIN

Alignment Scores:
Pred. No.:              9.7e-05      Length:         787
Score:                  197.50       Matches:        48
Percent Similarity:     48.51%       Conservative:   17
Best Local Similarity:  35.82%       Mismatches:     64
Query Match:            2.67%        Indels:         5
DB:                     14           Gaps:           1

US-09-913-878A-2 (1-1402) x CB907399 (1-787)

Qy    1257 ArgValAsnGlnValTyrGluLysTrpCysAlaIleThrProGluAlaMetAspLysSer 1276
                 :::         ||||||:::||||||   |||   |||
Db     110 AAGGCCGAGGCTGTCTACCAAAAATGGCTGGACATACAACCGCCGCCCGAGCTGCTGATG 169

Qy    1277 GlyAlaAsnTyrAspSerLysValIleArgLeuLeuGluLeuSerPheLeuAlaAspArg 1296
                 :::   :::   |||               :::   |||   |||::::::
Db     170 AGCGGCGAGTTTCCCGCGCTAGTCCATGACGCAAACAAGGCTTCGGCGTTGAGTAAC--- 226

Qy    1297 GluMetAsnThrTrpAlaLeuLeuArgAlaSerThrAlaPheLysLeuTyrTyrHisLys 1316
                       |||   ||||||:::|||||||||   ||||||   |||:::|||
Db     227 ------------TGGCAGCTGCTGAAAGCTTCGACCACTTTCAAGAAGTATCACCACAGC 274

Qy    1317 SerProLysPheValTrpGlnMetAlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMet 1336
           :::|||||||||||||||   |||||||||:::|||||||||||||||||||||
Db     275 GCGCCCAAGTTTGTCTGGAACATGGCCGGAAAGCAGCTTGCCTACATCAAGGCGATGGCG 334

Qy    1337 ThrSerArgProGlyGluGlyAlaProAlaLeuMetThrAlaPheMetTyrAlaGlyLeu 1356
              :::      ||||||   :::      |||:::            |||:::      |||
```

```
Db        335 CGAAACGACGCTGGCGAGACGTCAAAGGTGCTTGTGATTCCCGAGATGTGGGGGTCCTG 394

Qy       1357 MetProAspLysLysPheThrLysGlnTyrValAlaArgLeuGluGlyAspGlySerGlu 1376
                 ||||||||||||                               :::   |||
Db        395 CGGCCGGACAAGAAGCTCATCACCTCGCTGGCGATACAGAGGGAGGCTGCGCGGGATTCG 454

Qy       1377 TyrProAspProGluValTyrGluValLeuGlyAspAspAsp 1390
                             :::   ||||||      ||||||
Db        455 GAGAGTGCACTGGCTTTGGAGGAGGTGTTTGAGTTTGACGAT 496

RESULT 5
BF096404
LOCUS       BF096404                577 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST360431 tomato nutrient deficient roots Lycopersicon esculentum
            cDNA clone cLEW1106 5' sequence similar to RNA-directed RNA
            polymerase (Arabidopsis thaliana), mRNA sequence.
ACCESSION   BF096404
VERSION     BF096404.1  GI:10902114
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (tomato)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   van der Hoeven,R.S., Garvin,D.F., Matern,A.L., Holt,I.E., Liang,F.,
            Upton,J., Hansen,T.S., Ronning,C.M., Craven,M.B., Bowman,C.L.,
            Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J.
            and Tanksley,S.D.
  TITLE     Generation of ESTs from tomato nutrient-deficient roots
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1..577
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA492"
                     /db_xref="taxon:4081"
                     /clone="cLEW1106"
                     /tissue_type="roots"
                     /dev_stage="5-6 weeks old"
                     /lab_host="SOLR"
                     /clone_lib="tomato nutrient deficient roots"
                     /note="Vector: pBluescriptSKmCUadapt; Site_1: 5' EcoRI;
                     Site_2: 3' XhoI; Roots were harvested from plants grown
                     under the following deficiencies/stresses: 10 mM Al, Zn,
                     P, K, Fe,N. mRNA was isolated from indivdual treatments.
                     Proportional aliquots of mRNA of each treatment were mixed
                     and used for library construction."
ORIGIN

Alignment Scores:
Pred. No.:               0.000174      Length:          577
Score:                   191.50        Matches:         43
Percent Similarity:      50.88%        Conservative:    44
Best Local Similarity:   25.15%        Mismatches:      71
Query Match:             2.59%         Indels:          13
DB:                      10            Gaps:            2

US-09-913-878A-2 (1-1402) x BF096404 (1-577)

Qy        874 ThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAsp 893
                |||          ::::::::::::||||||   |||:::   :::   :::|||:::
Db         51 ACAGCGGCTATAATGTTGAGTGCAGGGTTTAAACCTCAAAGTGAACCTCATTTGAGAGGG 110

Qy        894 IleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArg 913
                :::          ::::::     :::          |||::::::|||        :::
Db        111 ATGTTGGCTAGCATTAGAGCTGCTCAGCTTGGCGACCTCAGGAATAAGACAAGGATGTTT 170

Qy        914 ValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsn 933
              |||            ::::::             |||      |||   ||||||:::
Db        171 GTTACTTCAGGAAGGTGGTTGATGGCTGTTTGGATGAATTAGGTGAACTTGAGCAAGGC 230

Qy        934 GluValHisValGlyPheSerSer---------------------------------- 941
              :::      :::      ||||||
Db        231 CAATGCTTTATTCAAGTGTCAAGCCCTTCTTTGGAGACCTGTTTTGTTAAGCATGGTCCA 290

Qy        942 LysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAla 961
              :::|||   :::   :::::::::       ::::::            |||::::::|||
Db        291 GAGTTTTCTGAAATCAAGAAAAATCTTCAAGTAGTAAAGGGCCTT---GTTGTAATTGCA 347

Qy        962 ArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGlu 981
              ::::::|||         |||   |||::::::   :::   ||||||         |||
Db        348 AAGAACCCGTGTCTTCATCCCGGGGATGTGAGGATTCTGGAGGCTGTAGATGTTCCTGGT 407

Qy        982 LeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLys 1001
              ||||||   |||   |||   ::::::|||      |||||||||   |||   :::
Db        408 TTACACCATCTCTATGATTGTCTGGTCTTCCCTCAGAAAGGGGATAGGCCACATTCAAAT 467

Qy       1002 LysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIle 1021
              :::   ||||||   |||   |||||||||:::      :::|||   ||||||      :::
Db        468 GAAGCATCAGGGAGTGACCTTGATGGTGATCTCTACTTTGTGACTTGGGATGAAAATCTC 527

Qy       1022 ValAspGlyPheValAsnAlaGluMetProLeu 1032
              :::         :::   ||||||:::
Db        528 ATTCCACCCAGTAAGAAAAGCTGGATGCCAATG 560

RESULT 6
BC015736
LOCUS       BC015736               3758 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens transforming, acidic coiled-coil containing protein 2,
            mRNA (cDNA clone IMAGE:4865697), with apparent retained intron.
ACCESSION   BC015736
VERSION     BC015736.1  GI:18266910
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3758)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE  22388257
    PUBMED  12477932
REFERENCE   2  (bases 1 to 3758)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
```

```
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Genome Sequence Centre,
          BC Cancer Agency, Vancouver, BC, Canada
          info@bcgsc.bc.ca
          Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
          Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
          Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
          Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
          Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
          Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
          Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
          George Yang, Scott Zuyderduyn, Marco Marra.

          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAL Plate: 34 Row: i Column: 24
          This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1..3758
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4865697"
                     /tissue_type="Colon, adenocarcinoma"
                     /clone_lib="NIH_MGC_15"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
ORIGIN

Alignment Scores:
Pred. No.:              0.00516       Length:        3758
Score:                  188.50        Matches:       194
Percent Similarity:     34.72%        Conservative:  141
Best Local Similarity:  20.10%        Mismatches:    338
Query Match:            2.55%         Indels:        293
DB:                     11            Gaps:          42

US-09-913-878A-2 (1-1402) x BC015736 (1-3758)

Qy      82 GluAlaLysAlaAlaSerSerAsnTrpValProLysAlaHisAlaAspProAspThrLeu 101
           ::::::   ||:::|||:::            |||::::::|||
Db     374 CAAAGCTCCGCCAGCTCCACCCCCACCACCCCCGAAGTCAT------------------ 415

Qy     102 ProTrpSerLysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThr 121
                             ||||||      |||   ||||||      |||
Db     416 ------------------CCCAGA------ACCCGAGGTCAGCACACAGCC--------- 442

Qy     122 ValLeuLeuGluValLeuAsnArgPheMetProProProAsnAsnThrProGlyArgThr 141
                                                   |||||||||||||||
Db     443 ---------------------------------------------ACCCCCGGAAGAACC 457

Qy     142 PheGlyArgThrLeuSerGly------Pro-SerGlyLeuSerArgProThrSerThrAs 159
                          :::|||           |||  :::||||||         ||||||:::
Db     458 AGTAGAGACTGGGTTTCACCATGATGACCAGGCTGGTCTCAAACTCCTGACCTCAAGT-- 515

Qy     159 nThrLysArgLysAspGluProAlaAsn------------------------------ 168
                        |||   |||||||:::
Db     516 -------------GATCTGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGATGTGG 562

Qy     169 ----------ValThrPheAlaAspProProLysArg---SerLeuThrArgSerAlaTh 184
                     |||          |||   |||:::     |||:::         |||
Db     563 TTCTGAGACAGTCCCTGTCCCTGATGGCCCACGGAGCGACTCGGTGGAAGGAAGTCCCTT 622

Qy     184 rGlyProProIleHisGlyAlaAlaIleProLeuLysPheProAspProValAsnThrGl 204
                ||||||  |||          :::                  |||:::
Db     623 CCGTCCCCCGTCACACTCCTTCTCTGCCGTCTTCGATGAAGACAAGCCGATA---GCCAG 679

Qy     204 ySerLysArgProSerLeuGluSerGluAsnLeuAsn--------------------- 216
              |||          :::|||:::    :::|||:::
Db     680 CAGTGGGACTTACAACTTGGACTTTGACAACATTGAGCTTGTGGATACCTTTCAGACCTT 739

Qy     217 ------------GlnCysThrLysArgAlaLysGlyLysLeuSerAspAsnValAlaAl 232
                                   |||       :::||||||:::          |||
Db     740 GGAGCCTCGTGCCTCAGACGCTAAGAATCAGGAGGGCAAAGTG------AACACACGGAG 793

Qy     232 aAlaAlaAlaProProValProIleAla---SerAlaLeuAspLysValProThrArgAr 251
             :::            |||||||||:::      |||   |||   :::       ::: ::
Db     794 GAAGTCCACGGATTCCGTCCCCATCTCTAAGTCTACACTGTCCCGGTCGCTCAGCCTGCA 853

Qy     251 gHisAlaAsnThrArgAspProThrAlaThrGlyHisArgArgAlaAspGlnValAspSe 271
           :   ::::::                 :::::::::||||:::       |||       :::
Db     854 AGCCAGTGACTTTGATGGTGCTTCTTCCTCAGGCAATCCCGAGGCCGTGGCCCTTGCCCC 913

Qy     271 rPheAspThrSerGlnGlyThrSerTyrGlySerSerValPheSerAlaCysArg----- 289
                       |||    |||:::|||         ||||||               :::
Db     914 AGATGCATATAGCACGGGTTCCAGCAGTGCTTCTAGTACCCTTAAGCGAACTAAAAAACC 973

Qy     290 ---------------HisAsnGlnSerThrThrGlnSerSerPheGluAlaProProSe 304
                          :::    ||||||:::                  |||    ||||||
Db     974 GAGGCCGCCTTCCTTAAAAAAGAAACAGACCACCAAGAAACCCACAGAGACCCCCCCAGT 1033

Qy     304 rGlnProArgGluLysArgProValAspAlaThrValPheGluAlaGlyHisLeuIleGl 324
             :::     ::::::     |||   :::    ::::::        :::|||    :::
Db    1034 GAAGGAGACGCAACAGGAGCCAGATGAAGAGAGCCTTGTCCCCAGTGGGGAGAATCTAGC 1093

Qy     324 uSerProSerLysGlyArgThrThrLysSerHis------------------IleAspAs 338
            |||      :::|||       :::     |||:::                  ::::::
Db    1094 ATCTGAGACGAAAACGGAATCTGCCAAGACGGAAGGTCCTAGCCCAGCCTTATTGGAGGA 1153

Qy     338 nGlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThrTyrTyrGluSerPh 358
               ||||||
Db    1154 GACGCCCCTTGAG----------------------------------------------- 1166

Qy     358 eProSerSerGlyGlyGluGlyAlaIleProGluProSerArgSerAsnGlyLeuAlaAr 378
             |||::::::|||      :::    |||    |||     |||     |||
Db    1167 -CCCGCTGCGGGGCCCAAAGCTGCCTGCCCTCTGGACTCAGAGAGTGTAGAAGGGGTTGT 1225

Qy     378 gSerGluGluSerAlaArgSerGlnValGlnValHisAlaProValValAlaAlaArgLe 398
                     |||         :::||||||              ||||||           ||
Db    1226 CCCCCCGGCTTCTGGAGGTGGCAGAGTGCAGAACTCACCCCCTGTCGGGAGGAAAACGCT 1285

Qy     398 uArgAsnIleTrpProLysPheProLysTrpLeuHisGluAlaProLeuAlaValAlaTr 418
           |                                                ||||||     |||
Db    1286 G----------------------------------------CCTCTTACCACGGCCCC 1303

Qy     418 pGluValThrArgLeuPheMetHisCysLysValAspLeuGluAsp-------------- 433
            |||         :::                           ||||||
Db    1304 GGAGGCAGGGGAGGTAACCCCATCGGATAGCGGGGGGCAAGAGGACTCTCCAGCCAAAGG 1363

Qy     434 -GluSerLeuGlyLeuLysTyrAsp---------------ProSerTrpSerThrAlaAr 448
             |||:::     |||:::::::|||                    ||||||        ::
Db    1364 GCACTCCGTAAGGCTGGAGTTTGACTATTCTGAGGACAAGAGTAGTTGGGACAACCAGCA 1423

Qy     448 gAsp-------------------------------------------------------- 449
           ::::
Db    1424 GGAAAACCCCCCTCCTACCAAAAAGATAGGCAAAAAGCCAGTTGCCAAAATGCCCCTGAG 1483

Qy     450 -ValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArgGlyLysProPhePr 469
                :::    ||||||       ::::||||||                   |||    ||
Db    1484 GAGGCCAAAGATGAAAAAGACACCCGAGAAACTTGACAACACTCCTGCCTCACCT---CC 1540

Qy     469 oGluLysPro------ProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSe 487
           |     |||          |||||||||:::   :::         |||      |||:::
Db    1541 CAGATCCCCTGCTGAACCCAATGACATCCCATTGCTAAAGGTACTTACACCTTTGAT-- 1598

Qy     487 rLysGlySerAlaValValLeuSerAlaValLeuAspTyr------AsnProAspAsnSe 505
                                             :::|||       :::||||:::    ::
Db    1599 ------------------------------ATTGACAAGTGGGATGACCCCAATTTTAA 1627

Qy     505 rProThrAlaProLeuTyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysAr 525
            :|||     :::              ::::::      |||    :::      :::|||
Db    1628 CCCTTTTTCTTCCACCTCAAAAATGCAGGAGTCTCCCAAACTGCCCCAACAA-------- 1679
```

```
Qy        525 gLeuThrArgArgPheGlyProAspArgPhePheGluIleLeuIlePro---------- 541
Db       1680 ----TCATACAACTTTGACCCAGACACCTGTGATGAGTCCGTTGACCCCTTTAAGACATC 1735

Qy        542 -SerProThrSerThrSerProSerValProProValValSerLysGlnProGlyAlaVa 561
Db       1736 CTCTAAGACCCCCAGCTCACCTTCTAAATCCCCAGCCTCCTTTGAGATCCCAGCCAGTGC 1795

Qy        561 lGluGluVal-IleGlnTrpLeuThrMetGlyGlnHisSerLeuValGlyArgGlnTrpA 581
Db       1796 TATGGAAGCCAATGGAGTGGACGGGGATGGGCTAAACAAGCCCGCCAAGAAGAAGAAGAC 1855

Qy        581 rgAla---------------PhePheAlaLysAspAlaGlyTyrArgLysProLeuArgG 596
Db       1856 GCCC-CTAAAGACTGACACATTTAGGGTGAAAAAGTCGCCAAAACGGTCTCCTCTCTCTG 1914

Qy        596 luPheGlnLeuArgAlaGluAspPro-----------------------LysProIleI 608
Db       1915 AT------CCACCTTCCCAGGACCCCACCCCAGCTGCTACACCAGAAACACCACCAGTGA 1968

Qy        608 leLysGluArgValHisPhePheAlaGluThrGlyIle--------------------- 620
Db       1969 TCTCTGCGGTGGTCCACGCCACAGATGAGGAAAAGCTGGCGGTCACCAACCAGAAGTGGA 2028

Qy        620 ------------------------------------------------------------ 620
Db       2029 CGTGCATGACAGTGGACCTAGAGGCTGACAAACAGGACTACCCGCAGCCCTCGGACCTGT 2088

Qy        621 --ThrPheArgProAspValPheLysThrArgSerValValProAlaGluGluProValG 640
Db       2089 CCACCTTT---------GTAAACGAGACCAAATTCAGTTCACCCACTGAGGAG---TTGG 2136

Qy        640 luGlnArgThrGluPheLysValSerGlnMet---------------------------- 650
Db       2137 ATTACAGAAACTCCTATGAAATTGAATATATGGAGAAAATTGGCTCCTCCTTACCTCAGG 2196

Qy        651 --------------LeuAspTrpLeuLeuGlnLeuAspAsnAsnThrTrpGlnProHisL 666
Db       2197 ACGACGATGCCCCGAAGAAGCAGGCCTTGTACCTTATGTTTGACACTTCTCAGGAGAGCC 2256

Qy        666 euLysLeuPheSerArgIleGlnLeuGlyLeuSer------------------------ 677
Db       2257 CTGTCAAGTCATCTCCCGTCCGCATGTCAGAGTCCCCGACGCCGTGTTCAGGGTCAAGTT 2316

Qy        678 -----LysThrTyrAlaIleMetThr---------------------------------- 684
Db       2317 TTGAAGAGACTGAAGCCCTTGTGAACACTGCTGCGAAAAACCAGCATCCTGTCCCACGAG 2376

Qy        685 --LeuGluProHisGlnIleArgHisHisLysThrAspLeuLeuSerProSerGlyThrG 704
Db       2377 GACTGGCCCCTAACCAAGAGTCACAC------------TTGCAGGTGCCAGAGAAATCCT 2424

Qy        704 lyGluValMetAsnAspGlyValGlyArgMetSerArgSerValAlaLysArgIleArgA 724
Db       2425 CCCAGAAGGAGCTGGAGGCCATGGGCTTGGGCACCCCTTCAGAAGCGATTGAAATTAGAG 2484

Qy        724 spValLeuGlyLeuGlyAspVal---------ProSerAlaValGlnGlyArgPheGlyS 741
Db       2485 AGGCTGCTCACCCAACAGACGTCTCCATCTCCAAAACAGCCTTGTACTCCCGCATCGGGA 2544

Qy        741 erAlaLys---------------GlyMetTrpVal------IleAspValAspAspThrG 754
Db       2545 CCGCTGAGGTGGAGAAACCTGCAGGCCTTCTGTTCCAGCAGCCCGACCTGGACTCTGCCC 2604

Qy        754 lyAspGluAspTrpIleGluThrTyrProSerGlnArg---------LysTrpGluCysA 771
Db       2605 TCCAGATCGCCAGAGCAGAGATCATAACCAAGGAGAGAGAGGTCTCAGAATGGAAAGATA 2664

Qy        771 spPheValAspLysHisGlnArgThrLeuGluValArgSerValAlaSerGluLeuLysS 791
Db       2665 AATATGAAGAAAGCAGGCGGGAAGTGATGGAAATGAGGAAAATAGTGGCCGAGTATGAGA 2724

Qy        791 erAlaGlyLeuAsnLeuGlnLeuLeuProValLeuGluAspArgAlaArgAspLys---- 809
Db       2725 AGACC---------------ATCGCTCAGATGATAGAGGACGAACAGAGAGAGAAGTCAG 2769

Qy        810 ----------------------------ValLysMetArgGlnAlaIleGlyAsp---- 818
Db       2770 TCTCCCACCAGACGGTGCAGCAGCTGGTTCTGGAGAAGGAGCAAGCCCTGGCCGACCTGA 2829

Qy        819 -----------ArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAlaL 835
Db       2830 ACTCCGTGGAGAAGTCTCTGGCCGACCTCTTCAGAAGATATGAGAAGATGAAGGAGGTCC 2889

Qy        835 eu---------AsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerA 852
Db       2890 TAGAAGGCTTCCGCAAGAATGAAGAGGTGTTGAAGAGATGTGCGCAGGAGTACCTGTCCC 2949

Qy        852 rgAlaThrArg 855
Db       2950 GGGTGAAGAAG 2960
```

RESULT 7
CF883391

```
LOCUS       CF883391                  737 bp    mRNA    linear   EST 31-OCT-2003
DEFINITION  tric088xh09.b1 T.reesei mycelial culture, Version 6 October 2003
            Hypocrea jecorina cDNA clone tric088xh09, mRNA sequence.
ACCESSION   CF883391
VERSION     CF883391.1  GI:38138073
KEYWORDS    EST.
SOURCE      Hypocrea jecorina (anamorph: Trichoderma reesei)
  ORGANISM  Hypocrea jecorina
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,
            Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and
            Dean,R.A.
  TITLE     Analysis of the protein processing and secretion pathways in a
            Trichoderma reesei EST dataset
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Ralph A. Dean
            Fungal Genomics Laboratory
            North Carolina State University
            Campus Box 7251, Raleigh, NC 27695, USA
            Tel: 919-513-0020
            Fax: 919-513-0024
            Email: ralph_dean@ncsu.edu
            Seq primer: LT-F1 primer.
FEATURES             Location/Qualifiers
     source          1..737
                     /organism="Hypocrea jecorina"
                     /mol_type="mRNA"
                     /strain="QM6a"
                     /db_xref="taxon:51453"
                     /clone="tric088xh09"
                     /dev_stage="mycelia"
                     /clone_lib="T.reesei mycelial culture, Version 6 October
                     2003"
                     /note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial
                     culture grown from 24 hrs to 6 days with varying Carbon
                     and Nitrogen sources and concentrations."
ORIGIN

Alignment Scores:
Pred. No.:              0.000611      Length:        737
Score:                  186.50        Matches:       50
Percent Similarity:     46.58%        Conservative:  18
Best Local Similarity:  34.25%        Mismatches:    71
Query Match:            2.52%         Indels:        7
DB:                     14            Gaps:          2

US-09-913-878A-2 (1-1402) x CF883391 (1-737)

Qy       1247 MetArgAsp------SerLysAspProTyrProValArgValAsnGlnValTyrGluLys 1264
```

```
Db          80 ATGCGCGACGGCATTGAAAACTCAAAGTTTGCCAACAAGGCCGAGGCTGTCTACCAAAAA 139

Qy        1265 TrpCysAlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysVal 1284
                |||      |||   |||                      :::   :::   |||
Db         140 TGGCTGGACATACAACCGCCGCCCGAGCTGCTGATGAGCGGCGAGTTTCCCGCGCTAGTC 199

Qy        1285 IleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrpAlaLeuLeu 1304
                           :::   |||   |||::::::                    |||||
Db         200 CATGACGCAAACAAGGCTTCGGCGTTGAGTAAC---------------TGTTTTTTGCTG 244

Qy        1305 ArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPheValTrpGlnMet 1324
               :::|||||||||   ||||||   |||:::|||   :::||||||||||||||   |||
Db         245 AAAGCTTCGACCACTTTCAAGAAGTATCACCACAGCGCGCCCAAGTTTGTCTGGAACATG 304

Qy        1325 AlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArgProGlyGluGlyAla 1344
               ||||||:::|||||||||||||||||||||         :::      |||||   :::
Db         305 GCCGGAAAGCAGCTTGCCTACATCAAGGCGATGGCGCGAAACGACGCTGGCGAGACGTCA 364

Qy        1345 ProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLysLysPheThrLys 1364
                        |||:::         |||:::      |||   |||||||||||
Db         365 AAGGTGCTTGTGATTCCCGAGATGTGGGGGTCCTGCGGCCGGACAAGAAGCTCATCACC 424

Qy        1365 GlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAspProGluValTyrGlu 1384
                           :::   |||                         :::   |||
Db         425 TCGCTGGCGATACAGAGGGAGGCTGCGCGGGATTCGGAGAGTGCACTGGCTTTGGAGGAG 484

Qy        1385 ValLeuGlyAspAspAsp 1390
               |||         ||||||
Db         485 GTGTTTGAGTTTGACGAT 502
```

```
RESULT 8
CB909520
LOCUS       CB909520                 776 bp    mRNA    linear   EST 02-JUL-2003
DEFINITION  tric088xh09 T.reesei mycelial culture, Version 3 april Hypocrea
            jecorina cDNA clone tric088xh09, mRNA sequence.
ACCESSION   CB909520
VERSION     CB909520.1  GI:30124181
KEYWORDS    EST.
SOURCE      Hypocrea jecorina (anamorph: Trichoderma reesei)
  ORGANISM  Hypocrea jecorina
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE   1  (bases 1 to 776)
  AUTHORS   Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
            Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
            Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
            Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
  TITLE     Transcriptional regulation of biomass-degrading enzymes in the
            filamentous fungus Trichoderma reesei
  JOURNAL   J. Biol. Chem. 278 (34), 31988-31997 (2003)
  MEDLINE   22803314
   PUBMED   12788920
COMMENT     Contact: Pamela K. Foreman
            Genencor Intl.
            925 Page Mill Road, Palo Alto, CA 94304, USA
            Tel: (650) 846-7635
            Fax: (650) 621-7817
            Email: Pforeman@genencor.com
            Seq primer: LT-F1 primer.
FEATURES             Location/Qualifiers
     source          1. .776
                     /organism="Hypocrea jecorina"
                     /mol_type="mRNA"
                     /strain="QM6a"
                     /db_xref="taxon:51453"
                     /clone="tric088xh09"
                     /dev_stage="mycelia"
                     /clone_lib="T.reesei mycelial culture, Version 3 april"
                     /note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial
                     culture grown from 24 hrs to 6 days with varying Carbon
                     and Nitrogen sources and concentrations."
ORIGIN
```

```
Alignment Scores:
Pred. No.:                0.000661      Length:         776
Score:                    186.50        Matches:        50
Percent Similarity:       46.58%        Conservative:   18
Best Local Similarity:    34.25%        Mismatches:     71
Query Match:              2.52%         Indels:         7
DB:                       14            Gaps:           2

US-09-913-878A-2 (1-1402) x CB909520 (1-776)

Qy        1247 MetArgAsp------SerLysAspProTyrProValArgValAsnGlnValTyrGluLys 1264
               |||||||||                     :::   :::         ||||||:::|||
Db          80 ATGCGCGACGGCATTGAAAACTCAAAGTTTGCCAACAAGGCCGAGGCTGTCTACCAAAAA 139

Qy        1265 TrpCysAlaIleThrProGluAlaMetAspLysSerGlyAlaAsnTyrAspSerLysVal 1284
                |||      |||   |||                      :::   :::   |||
Db         140 TGGCTGGACATACAACCGCCGCCCGAGCTGCTGATGAGCGGCGAGTTTCCCGCGCTAGTC 199

Qy        1285 IleArgLeuLeuGluLeuSerPheLeuAlaAspArgGluMetAsnThrTrpAlaLeuLeu 1304
                           :::   |||   |||::::::                    |||||
Db         200 CATGACGCAAACAAGGCTTCGGCGTTGAGTAAC---------------TGTTTTTTGCTG 244

Qy        1305 ArgAlaSerThrAlaPheLysLeuTyrTyrHisLysSerProLysPheValTrpGlnMet 1324
               :::|||||||||   ||||||   |||:::|||   :::||||||||||||||   |||
Db         245 AAAGCTTCGACCACTTTCAAGAAGTATCACCACAGCGCGCCCAAGTTTGTCTGGAACATG 304

Qy        1325 AlaGlyArgGlnLeuAlaTyrIleLysAlaGlnMetThrSerArgProGlyGluGlyAla 1344
               ||||||:::|||||||||||||||||||||         :::      |||||   :::
Db         305 GCCGGAAAGCAGCTTGCCTACATCAAGGCGATGGCGCGAAACGACGCTGGCGAGACGTCA 364

Qy        1345 ProAlaLeuMetThrAlaPheMetTyrAlaGlyLeuMetProAspLysLysPheThrLys 1364
                        |||:::         |||:::      |||   |||||||||||
Db         365 AAGGTGCTTGTGATTCCCGAGATGTGGGGGTCCTGCGGCCGGACAAGAAGCTCATCACC 424

Qy        1365 GlnTyrValAlaArgLeuGluGlyAspGlySerGluTyrProAspProGluValTyrGlu 1384
                           :::   |||                         :::   |||
Db         425 TCGCTGGCGATACAGAGGGAGGCTGCGCGGGATTCGGAGAGTGCACTGGCTTTGGAGGAG 484

Qy        1385 ValLeuGlyAspAspAsp 1390
               |||         ||||||
Db         485 GTGTTTGAGTTTGACGAT 502
```

```
RESULT 9
CG445102/c
LOCUS       CG445102                 954 bp    DNA     linear   GSS 17-SEP-2003
DEFINITION  OG8AH57TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0850J18,
            genomic survey sequence.
ACCESSION   CG445102
VERSION     CG445102.1  GI:34827337
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 954)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .954
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
```

```
                  /strain="B73"
                  /db_xref="taxon:4577"
                  /clone="ZMMBMa0850J18"
                  /clone_lib="ZM_0.7_1.5_KB"
                  /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                  methylation filtered genomic DNA library"
ORIGIN

Alignment Scores:
Pred. No.:               0.00129        Length:         954
Score:                   184.50         Matches:        61
Percent Similarity:      41.51%         Conservative:   49
Best Local Similarity:   23.02%         Mismatches:     91
Query Match:             2.49%          Indels:         65
DB:                      29             Gaps:           8

US-09-913-878A-2 (1-1402) x CG445102 (1-954)

Qy        871 GlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyr 890
Db        818 CAAGGAAGCACCGCAGGGCTGATGCTGGGCGCTGGCTTCGCTCCTGCGACTGAGCCACAC 759

Qy        891 LeuGlnAspIleAla-TrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLe 910
Db        758 CTGAGAGCGATGCTACTGGCCATCTGCACCTCGCAGATGCAGGGCCTTCTCAACAAAACG 699

Qy        910 uAsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLe 930
Db        698 TGG-ATCTTTGTGCCCAAGGGAAGGTGGCTGATGGGCTGTCTCGACGAGTTTGGGATCCT 640

Qy        930 uGluGluAsnGluVal--------------------HisValGly------------- 938
Db        639 TGAGCAGGGGCAGTGCTTTCTCCGGGTCTCGACGCCATCAGTCGGCAGCCGCTGCGTGAA 580

Qy        939 -------------PheSerSerLysPheArgAspGluGluGluSerPheThrLeuLeuSe 954
Db        579 CCGTTCTCCGATATTTCCCTCGGAGTATAAGGGCGATGCCGATGCACAGGTGATTACTGG 520

Qy        954 rAspCysAspValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgVa 974
Db        519 T------ACGGTTGTGATGGCGAAGAACCCATGCCTTCATCCAGGGGACGTCCGTATCCT 466

Qy        974 lArgAlaValPheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLy 994
Db        465 TGAAGCTGTGGATGTCCCTGATCTGTATCACCTTGTCGACTGCTTCGTCTTCCCCAAGAA 406

Qy        994 sGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTr 1014
Db        405 AGGTGAGAGGCCACACGCCGACGAAGCTTCCGGGAGCGACCTCGACGGCGACGTCTACTT 346

Qy       1014 pValCysTrpAspProGluIleValAspGlyPheValAsnAlaGluMetProLeuGlu-- 1033
Db        345 CGTGACGTGGGACGAGAGCCTTGTACCTCCCGGGAGGAAGAGCTGCACCCCGATGGACTA 286

Qy       1034 ----ProAspLeuSerArgTyrLeuLysLysAspLysThrThrPheLysGlnLeuMetAl 1052
Db        285 CTCCCCGGCTAAAACAAAGCAACTTCCACGCGAC-------------------------- 252

Qy       1052 aSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSe 1072
Db        251 ------------------------------------GTACACCAACAAGTAAGTACACA 229

Qy       1072 rPheHisPheAlaLeuGlnProAsnPhe---LeuGlyMetCysThrAsn----------- 1087
Db        228 ATACCACTTTGCT------TCTTCCTTTGGTTTATCTTGCTGCCATAATACATCAAGTTA 175

Qy       1088 ---------------------------------------------TyrLysGluArgLe 1092
Db        174 ATGGATTTCTCGCGTGCTCGTGTTGTAACAGGATACCATTGATTTCTACTTGGAGAGCAT 115

Qy       1092 uCysTyrIleAsn 1096
Db        114 GGTATACGACAAC 102
```

```
RESULT 10
CA257764
LOCUS       CA257764                 641 bp    mRNA    linear   EST 26-SEP-2003
DEFINITION  SCCCFL1100F01.g FL1 Saccharum officinarum cDNA clone SCCCFL1100F01
            5', mRNA sequence.
ACCESSION   CA257764
VERSION     CA257764.1  GI:35945014
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 100  row: F  column: 01
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1..641
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCCCFL1100F01"
                     /lab_host="DH10B"
                     /clone_lib="FL1"
                     /note="Organ: Inflorescence at begining of development
                     (1cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI;
                     An unidirectional cDNA library generated from
                     [Inflorescence at begining  of development (1cm-long)].
                     cDNA was prepared from polyA+ mRNA using  SuperScript
                     Plasmid System Kit (Invitrogen). The double-strand cDNAs
                     were  fractionated in a sepharose CL-2B 40cm-columns and
                     fragments sizing  between 0.8 and 1.5 Kb were
                     directionally cloned into the vector. Details  of each
                     source of RNA and library construction can be obtained at
                     http://sucest.lad.ic.unicamp.br/public"
ORIGIN

Alignment Scores:
Pred. No.:               0.000838       Length:         641
Score:                   183.50         Matches:        44
Percent Similarity:      50.30%         Conservative:   39
Best Local Similarity:   26.67%         Mismatches:     67
Query Match:             2.48%          Indels:         15
DB:                      14             Gaps:           3

US-09-913-878A-2 (1-1402) x CA257764 (1-641)

Qy        871 GlnGluGluThrLeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyr 890
Db         25 CAAGGAAACACTGCAGCACTGATGTTGAGTGCTGGCATTTCACCTGGAACTGAACCACAC 84

Qy        891 LeuGlnAspIleAlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeu 910
Db         85 CTGAAAGCAATGCTGTTAGCTATAAGGTCCTCACAGCTACTAGGCCTTTTGGAGAAAACA 144

Qy        911 AsnIleArgValGlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeu 930
Db        145 AGGATTTTTGTGCCCAAGGGGAGGTGGTTGATGGGCTGCCTTGATGAACTTGGGATCCTT 204
```

```
Qy       931 GluGluAsnGluValHisValGlyPheSerSer-------------------------- 941
             |||:::   :::      :::      ||||||
Db       205 GAGCAAGGACAGTGCTTTATCCGGGCGTCATCTCCATCACTCAATAATTGTCTAGTAAAG 264

Qy       942 ---------LysPhe---ArgAspGluGluGluSerPheThrLeuLeuSerAspCysAsp 957
                      :::|||   :::       :::  :::    |||::::::
Db       265 CATGGACCAAGATTTTCCAAAGCAAACAAAAATGCAGAGACCATTGTGGGT------ACT 318

Qy       958 ValLeuValAlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaVal 977
             :::::::::|||::::::|||             |||    |||::::::   :::   ||||||
Db       319 ATTGTAATGGCAAAGAATCCATGTCTTCATCCAGGGGATGTCCGGATTCTTGAAGCTGTT 378

Qy       978 PheLysProGluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspVal 997
                      |||||||||||    |||    |||    ::::::|||       ||||||||
Db       379 GATGTGCCTGAACTGCATCACCTTGTTGATTGCTTGGTCTTCCCTAAGAAAGGTGATAGG 438

Qy       998 ProLeuAlaLysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrp 1017
             |||   |||    :::    ||||||    |||    |||||||||:::   :::|||    |||
Db       439 CCACATGCCAATGAAGCATCTGGGAGTGATCTTGATGGGGATCTCTACTTTGTGACATGG 498

Qy      1018 AspProGluIleVal 1022
              |||     ::::::
Db       499 GATGAAAACCTTATA 513
```

RESULT 11
BI246074
LOCUS BI246074 582 bp mRNA linear EST 17-JUL-2001
DEFINITION IP1_68_D10.b1_A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BI246074
VERSION BI246074.1 GI:14824093
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 582)
AUTHORS Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and Pratt,L.H.
TITLE An EST database from Sorghum: developing preanthesis pannicles
JOURNAL Unpublished (2001)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyTMix or T7 sequencing primer, are presented as the reverse complement.
Seq primer: JEN REV
High quality sequence stop: 560
POLYA=No.
FEATURES Location/Qualifiers
source 1. .582
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone_lib="Immature pannicle 1 (IP1)"
/note="Organ: Developing preanthesis pannicles; Vector: pBluescript II SK(-) from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."
ORIGIN

```
Alignment Scores:
Pred. No.:              0.00297        Length:          582
Score:                  175.50         Matches:         42
Percent Similarity:     49.69%         Conservative:    38
Best Local Similarity:  26.09%         Mismatches:      60
Query Match:            2.37%          Indels:          21
DB:                     12             Gaps:            3

US-09-913-878A-2 (1-1402) x BI246074 (1-582)

Qy       878 LeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIleAlaTrpAsp 897
             ::::::::::::|||        |||       :::    :::|||:::   :::
Db        17 ATGTTGAGTGCTGGCATTTCACCTGGAACTGAGCCACACCTGAAAGCAATGCTGTTAGCT 76

Qy       898 LeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSer 917
             ::::::    :::             |||      |||       |||   |||   :::
Db        77 ATAAGGTCCTCACAGCTACTGGGTCTTTTGGAGAAGACAAGGATTTTTGTGCCGAAGGGG 136

Qy       918 AlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisVal 937
                :::::             |||       |||:::||||||:::   :::      :::
Db       137 AGGTGGTTGATGGGCTGCCTTGATGAACTTGGGATCCTTGAGCAAGGACAGTGCTTTATC 196

Qy       938 ---------------------------------------------GlyPheSerSer 941
                                                          ||||||||
Db       197 CGGGCGTCATCTCCATCACTCAATAATTGTCTGGTAAAGCATGGACCAAGATTTTCCTCA 256

Qy       942 LysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuValAla 961
                   ::::::   |||          |||::::::          :::::::::|||
Db       257 GCAAACAAAAATGCAGAG---------ACCATTGTGGGT------ACTATCGTAATGGCA 301

Qy       962 ArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysProGlu 981
             ::::::|||          |||    |||::::::   :::   ||||||       ||||||
Db       302 AAGAATCCATGCCTTCATCCAGGGGATGTCCGGATCCTTGAAGCTGTTGATGTGCCTGAA 361

Qy       982 LeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAlaLys 1001
             ||||||   |||    |||    ::::::|||       ||||||:::   |||    |||
Db       362 CTGCATCACCTTGTTGATTGCTTGGTCTTCCCCAAGAAAGGTGAGAGGCCACATGCCAAT 421

Qy      1002 LysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIle 1021
             :::    ||||||    |||    |||||||||:::    :::|||    ||||||       :::
Db       422 GAAGCATCTGGGAGTGATCTTGATGGGGATCTCTACTTTGTGACATGGGATGAAAACCTT 481

Qy      1022 Val 1022
             :::
Db       482 ATA 484
```

RESULT 12
BU965311
LOCUS BU965311 622 bp mRNA linear EST 21-OCT-2002
DEFINITION sat08g07.y1 Gm-c1036 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-c1036-13357 5' similar to TR:Q9ZR58 Q9ZR58 RNA-DIRECTED RNA POLYMERASE ;, mRNA sequence.
ACCESSION BU965311
VERSION BU965311.1 GI:24206058
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 622)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

```
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact: ccu@resgen.com web site:
            www.resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 445.
FEATURES             Location/Qualifiers
     source          1..622
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1036-13357"
                     /tissue_type="somatic embryos cultured on MSD 20"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1036"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
                     cDNA library was constructed from mRNA isolated from
                     somatic  embryos (age ranging from 2 months to 9 months)
                     cultured on MSD 20. The library was prepared using the
                     Life Technologies pSuperScript cDNA library construction
                     kit. Complementary DNA was synthesized from mRNA using a
                     poly (dT) sequence with a NotI restrictions site. SalI
                     linkers adapters were ligated to the blunt-ended cDNA
                     fragments followed by NotI digestion. The cDNA fragments
                     were directionally cloned into the NotI-SalI restriction
                     site of the pSPORT1 vector. The ligated cDNA fragments
                     were transformed into E.coli ElectroMax DH10B host cells.
                     This library was constructed in the laboratory of Dr. Lila
                     Vodkin by Anu Khanna at the University of Illinois at
                     Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"
ORIGIN

Alignment Scores:
Pred. No.:              0.0051       Length:         622
Score:                  173.00       Matches:        44
Percent Similarity:     50.00%       Conservative:   34
Best Local Similarity:  28.21%       Mismatches:     70
Query Match:            2.34%        Indels:         8
DB:                     13           Gaps:           2

US-09-913-878A-2 (1-1402) x BU965311 (1-622)

Qy        875 LeuAsnPheLeuMetAsnSerGlyPheAspProLysLysGlnLysTyrLeuGlnAspIle 894
              |||      ::::::      |||:::   |||   ::::::   :::|||     :::
Db        149 CTGAAGGAGATGCTCATTTGTGGATACAAGCCTAATGAAGAACCATTCCTTTCAATGATG 208

Qy        895 AlaTrpAspLeuGlnLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgVal 914
                       :::      |||            |||:::   |||      |||      :::
Db        209 CTTCAAACATTTAGGGCATCAAAACTTTTGGAATTGCGACTTAAATCTAGGATCTTTATT 268

Qy        915 GlyArgSerAlaTyrIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGlu 934
                 :::      :::            |||         ||||||         :::
Db        269 CCAAAAGGAAGAGCAATGATGGGATGTCTAGATGAAACTAGAACCCTAGAATATGGTCAA 328

Qy        935 ValHisValGlyPheSerSer------------LysPheArgAspGluGluGluSerPhe 950
              |||   |||   ||||||:::                     |||:::      |||:::
Db        329 GTATTTGTTCAGTTCTCTAACAATAGGCTGCAGAATCTATCTGATGATTTTTTTCATAT 388

Qy        951 ThrLeu------------LeuSerAspCysAspValLeuValAlaArgSerProAlaHis 966
                 |||                  :::      |||:::||||||:::::::||| 
Db        389 GATTTGCCAAAGAATTATATGGTTAAAGGTAAGGTAGTTGTAGCAAAAAACCCCTGCTTG 448

Qy        967 PheProSerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLys 986
                 |||   |||::::::   ::::::|||||         |||:::|||:::   :::
Db        449 CACCCAGGTGATGTGCGTGTTTTACAAGCTGTGGATGTGCCAGATTTGTACCACATGGTG 508

Qy        987 AspValIleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGly 1006
              |||   ::::::|||      |||||         |||         :::      |||||
Db        509 GACTGTGTTGTTTTCCCTCAAAAAGGACCAAGACCTCATCCAAATGAGTGTTCGGGAAGT 568

Qy       1007 AspTyrAspGlyAspMetAlaTrpValCysTrpAspProGluIleVal 1022
              |||   |||||||||:::   :::|||||||||||||   |||::::::
```

```
Db        569 GATCTGGATGGAGATATCTACTTTGTTTGTTGGGACCATGAAATGATT 616

RESULT 13
BF275078
LOCUS       BF275078               761 bp    mRNA    linear   EST 07-MAR-2001
DEFINITION  GA__Eb0023B06f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Eb0023B06f, mRNA sequence.
ACCESSION   BF275078
VERSION     BF275078.1  GI:11206148
KEYWORDS    EST.
SOURCE      Gossypium arboreum
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 761)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J.,
            Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 751.
FEATURES             Location/Qualifiers
     source          1..761
                     /organism="Gossypium arboreum"
                     /mol_type="mRNA"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Eb0023B06f"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Alignment Scores:
Pred. No.:              0.00828      Length:         761
Score:                  172.00       Matches:        66
Percent Similarity:     40.14%       Conservative:   46
Best Local Similarity:  23.66%       Mismatches:     102
Query Match:            2.33%        Indels:         65
DB:                     10           Gaps:           9

US-09-913-878A-2 (1-1402) x BF275078 (1-761)

Qy        941 SerLysPheArgAspGluGluGluSerPheThrLeuLeuSerAspCysAspValLeuVal 960
              |||:::|||   :::   ::::::::::::   ::::::               |||::::::
Db         36 TCTAGGTTTGCTGAAACAAAGAAAAATTGGGAAGTAATTAAGGGGCTG---GTTGTCATA 92

Qy        961 AlaArgSerProAlaHisPheProSerAspIleGlnArgValArgAlaValPheLysPro 980
              |||::::::|||               |||   ||||||:::   :::   ||||||      |||
Db         93 GCTAAGAATCCTTGTCTTCACCCTGGAGATATAAGAATTCTAGAAGCAGTTGATGCCCCT 152

Qy        981 GluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAla 1000
                 ||||||   |||   |||   ::::::|||         ||||||:::   |||
Db        153 GGTTTACACCATTTGTATGACTGTCTGGTCTTCCCTCAAAAGGGTGAGAGACCCCATACG 212

Qy       1001 LysLysLeuSerGlyGlyAspTyrAspGlyAspMetAlaTrpValCysTrpAspProGlu 1020
                 :::   ||||||   |||   |||||||||:::   :::|||   |||:::
Db        213 AATGAAGCTTCTGGAAGTGATCTTGATGGGGACCTTTATTTGTCACATGGGAAGAGGCT 272

Qy       1021 IleValAspGlyPheValAsnAlaGluMetProLeu------GluProAspLeuSerArg 1038
              ::::::               :::      |||:::      :::||||||      |||
```

```
Db      273 CTTATCCCCCCTAGCAAGAAGAGCTCACAACCCATGCAGTATGACCCTGATGAGCCTCGA 332

Qy     1039 TyrLeuLysLysAspLysThr-----------------ThrPheLysGlnLeuMetAla 1052

Db      333 GAGTTGAATCGCCCAGTCACTCATAAGGACATAATTGAGTTTTTTCAAAAAACATGGTG 392

Qy     1053 SerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyrAspMetIleGlnLysSer 1072

Db      393 AATGAGCACCTGGGAAGTATCTGCAATGCACATGTGGTTCATTCTGACCTTAGTGAA--- 449

Qy     1073 PheHisPheAlaLeuGlnProAsnPheLeuGlyMetCysThrAsnTyrLysGluArgLeu 1092

Db      449 ------------------------------------------------------------ 449

Qy     1093 CysTyrIleAsnAsnSerValSerAsnLysProAlaIleIleLeuSerSerLeuValGly 1112

Db      450 ------------CATGGTGCTTCAGATGAGAAATGTATACATCTCGCAGAGTTAGCGGCT 497

Qy     1113 AsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAlaGlnLeu 1132

Db      498 ATAGCTGTTGATTTTCCTAAAACCGGGAAGATTGTGTCGATGCCT------GCTCAATTA 551

Qy     1133 ArgArgGluLeuLeuGlyGlyAlaLeuSerLeuProAspProMet--------------- 1147

Db      552 AAACCAAAACTT-----------------TACCCAGATTTTATGGGTAAAGAGGAGTTC 593

Qy     1148 -------TyrLysSerAspSerTrpLeuGlyArgGlyGluProThrHisIleIleAspTy 1165

Db      594 CAGTCATTACAAGTCAAACAAAATTCTGGGAAGA------------------------- 627

Qy     1165 rLeuLysPheSerIleAlaArgProAlaIleAspLysGluLeuGluAlaPheHisAsnAl 1185

Db      628 ----CTATATCGTTATATCAGAGATGCTTATGACCAAGATGTCTCTGAATCTTCTGAGCT 683

Qy     1185 aMetLysAlaAlaLysAspThrGluAspGlyAlaHisPheTrpAspProAspLeu 1203

Db      684 CAATTTTGGTGCTAGTGACATC--------------ATTTATGATGCCGATCTT 723
```

```
RESULT 14
CG212111
LOCUS       CG212111                 934 bp    DNA     linear   GSS 22-AUG-2003
DEFINITION  OG0FJ51TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0705J06,
            genomic survey sequence.
ACCESSION   CG212111
VERSION     CG212111.1  GI:34111941
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 934)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OG0FJ51TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..934
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0705J06"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

Alignment Scores:
Pred. No.:              0.0135       Length:          934
Score:                  171.00       Matches:         67
Percent Similarity:     37.84%       Conservative:    45
Best Local Similarity:  22.64%       Mismatches:      84
Query Match:            2.31%        Indels:          100
DB:                     29           Gaps:            12

US-09-913-878A-2 (1-1402) x CG212111 (1-934)
```

```
Qy      925 AspPheTrpGlyValLeuGluGluAsnGluVal--------------------HisVal 937

Db       27 GACGAGTTTGGGATCCTTGAGCAGGGGCAGTGCTTTCTCCGGGTCTCGACGCCATCAGTC 86

Qy      938 Gly-------------------------PheSerSerLysPheArgAspGluGluGlu 948

Db       87 GGCAGCCGCTGCGTGAACCGTTCTCCGATATTTCCCTCGGAGTATAAGGGCGATGCCGAT 146

Qy      949 SerPheThrLeuLeuSerAspCysAspValLeuValAlaArgSerProAlaHisPhePro 968

Db      147 GCACAGGTGATTACTGGT------ACGGTTGTGATGGCGAAGAACCCATGCCTTCATCCA 200

Qy      969 SerAspIleGlnArgValArgAlaValPheLysProGluLeuHisSerLeuLysAspVal 988

Db      201 GGGGACGTCCGTATCCTTGAAGCTGTGGATGTCCCTGATCTGTATCACCTTGTCGACTGC 260

Qy      989 IleIlePheSerThrLysGlyAspValProLeuAlaLysLysLeuSerGlyGlyAspTyr 1008

Db      261 TTCGTCTTCCCCAAGAAAGGTGAGAGGCCACACGCCGACGAAGCTTCCGGGAGCGACCTC 320

Qy     1009 AspGlyAspMetAlaTrpValCysTrpAspProGluIleValAspGlyPheValAsnAla 1028

Db      321 GACGGCGACGTCTACTTCGTGACGTGGGACGAGAGCCTTGTACCTCCCGGGAGGAAGAGC 380

Qy     1029 GluMetProLeuGlu------ProAspLeuSerArgTyrLeuLysLysAspLysThrThr 1046

Db      381 TGCACCCCGATGGACTACTCCCCGGCTAAAACAAAGCAACTTCCACGCGAC--------- 431

Qy     1047 PheLysGlnLeuMetAlaSerHisGlyThrGlySerAlaAlaLysGluGlnThrThrTyr 1066

Db      432 ------------------------------------------------------GTACAC 437

Qy     1067 AspMetIleGlnLysSerPheHisPheAlaLeuGlnProAsnPhe---LeuGlyMetCys 1085

Db      438 CAACAAGTAAGTACACAATACCACTTTGCT------TCTTCCTTTGGTTTATCTTGCTGC 491

Qy     1086 ThrAsn------------------------------------------------------ 1087

Db      492 CATAATACATCAAGTTAATGGATTTCTCGCGTGCTCGTGTTGTAACAGGATACCATTGAT 551

Qy     1088 ---TyrLysGluArgLeuCysTyrIleAsn----------------------------- 1096

Db      552 TTCTACTTGGAGAGCATGGTATACGACAACCTTGGTCGAATATGCAACGCCCATGCTGCT 611

Qy     1097 -----------------AsnSerValSerAsnLysProAlaIleIleLeuSerSerLeu 1110

Db      612 CATGCTGATCGGAGCGACGACGGAGCGATGGATCCCAAGTGCATCGAGTTGGCTCGGCTC 671

Qy     1111 ValGlyAsnLeuValAspGlnSerLysGlnGlyIleValPheAsnGluAlaSerTrpAla 1130

Db      672 GCATCTATCGCCGTGGACTCCGCCAAGACCGGGGAAATTGTAAGA--------------- 716

Qy     1131 GlnLeuArgArgGluLeuLeuGlyGlyAlaLeuSer---------LeuProAspProMet 1147

Db      717 -----------------ATGCCACCAGCCCTTTCGCCAAAAGAGTACCCCGACTTCATG 758

Qy     1148 --------------------TyrLysSerAspSerTrpLeuGlyArg 1156
```

```
Db        759 GGGAAGGAGGACGCCATCTCTTACAGATCCGAAAAGATCCTTGGGAGG 806

RESULT 15
BC036553
LOCUS       BC036553                3267 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens cDNA clone IMAGE:5265665, containing frame-shift
            errors.
ACCESSION   BC036553
VERSION     BC036553.1  GI:23306459
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3267)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 3267)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 73 Row: m Column: 13
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1..3267
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5265665"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                 0.109        Length:         3267
Score:                     170.00       Matches:        226
Percent Similarity:        33.98%       Conservative:   143
Best Local Similarity:     20.81%       Mismatches:     413
Query Match:               2.30%        Indels:         306
DB:                        11           Gaps:           54

US-09-913-878A-2 (1-1402) x BC036553 (1-3267)

Qy         66 TrpArgLysAspAspSerLeuAsnGlnAlaGluAlaAsnPhePheIleGluAlaLysAla 85
              |||||:::   |||:::              ::::::          :::   |||||
Db         94 TGGAGACAGAGGGATGCGGCGGGAGGGCGGGACAGCCCA------GTCGCCGCCAAGGAG 147

Qy         86 Ala------------SerSerAsnTrpValProLysAlaHisAlaAspProAspThrLeu 101
              |||            |||:::      |||::::::          |||
Db        148 GCGAAGCGGGGTCGGGGTAGCTCACCCGCCCCCCGCAGTCCCCGCCCCCCG--------- 198

Qy        102 ProTrpSerLysGluProProArgAlaAlaThrAlaGlyGlnGlnTrpAlaLeuGlnThr 121
                    :::   ||||||||||||||
Db        199 ---------CAGTCCCCGCCCCGCGCGGCGCCGCGCACGCCT------------------ 231

Qy        122 ValLeuLeuGluValLeuAsnArgPheMetProProProAsnAsnThrPro--------- 138
                       ||||||:::   |||      |||         |||
Db        232 ---------GAAGTCATTGCCCGAGGCCCCGCCCCCTCGCGGCCCCAGCCCCGGCTCTGG 282

Qy        139 ---------------GlyArgThrPheGlyArgThrLeuSerGlyProSer--------- 150
                             |||||                 :::   |||
Db        283 CTCTGCCGGCGGAGGGGGCGGAGAGAACCCGCAGCTGCAACTTCGCCTGGAGCCTTTGAG 342

Qy        151 ---GlyLeuSerArgProThrSerThrAsnThrLysArg---LysAspGluProAlaAsn 168
                 |||||||||||   :::|||:::      :::|||            |||||
Db        343 CCTGGGCTCAGCCGGGGATCTAGCAGCCGTCACCGCCGAGTCGCCGCTTCCCCGGCGCGG 402

Qy        169 ValThr---------------------------------------------------Phe 171
                 :::
Db        403 GCCTCAGCCCAGCCGGCTTGGCCATGGCAGGCGCCGGCAGTGCGGCGAATCTGGCCCCTG 462

Qy        172 AlaAspProPro------LysArgSerLeuThrArgSerAlaThrGlyProPro------ 187
                    ||||||      ::::::   ||||||                ||||||
Db        463 CGCTCCCCGCCCGTAGCTGAGCAGCCGTTAACCAGCCATCCCGCCTTTCCTCCTGCCCGT 522

Qy        188 IleHisGlyAlaAlaIleProLeuLysPheProAspProValAsnThrGlySerLys--- 206
              :::      |||      |||   ||||||||   |||            |||   :::
Db        523 GTCGCCGAGGCCCACGCTCCGCGCAAGTTTCCGCCGCCGAA-GAAGCAGGAGGACGGACG 581

Qy        207 ------------ArgProSerLeuGluSerGluAsnLeuAsnGlnCysThrLysArgAla 222
                          ||||||               :::      :::|||
Db        582 TGAGGAGTCGGGCGTCCCTGGCCCACCCGGCCCTCTCGCCCGCGATGT------------ 629

Qy        223 LysGlyLysLeuSer------AspAsnValAlaAlaAlaAlaAlaProProValProIle 240
              |||      |||               ||||||         ||||||      |||
Db        630 ---GGGGGCGCTTCCTGGCTCCGGAGGCCAGCGGCCGGGATAGCCCCGGCGGAGCCCGCA 686

Qy        241 AlaSerAlaLeuAspLysVal--------------------------------------- 247
              ||||||            ::::::
Db        687 GCTTCCCTGCCGGCCCAGATTATTCATCAGCATGGCTACCTGCTAATGAATCATTGTGGC 746

Qy        248 ---ProThrArgArgHisAlaAsnThrArgAspProThrAlaThrGlyHisArgArgAla 266
                 |||         |||   |||   |||:::                |||   |||
Db        747 AGGCCACAACTGTTCCATCT-AACCATCGAAATAACCATATCAGGAGACACAGTATAGCT 805

Qy        267 AspGlnValAspSerPheAspThrSerGlnGlyThrSerTyrGlySerSerValPheSer 286
                       ||||||   ||||||   |||||||||         ||||||
Db        806 TCT------GACAGTGGAGACACTGGCATTGGAACTTCCTGTTCTGATAGCGTG------ 853

Qy        287 AlaCysArgHisAsnGlnSerThrThrGlnSerSerPheGluAlaProProSerGlnPro 306
                    :::   ||||||:::      :::               |||||||||
Db        854 ---------GAAGATCATTCAACTTCAAGTGGCACATTATCTTTTAAGCCTAGTCAATCA 904

Qy        307 ArgGluLysArgProValAspAlaThrValPheGluAlaGlyHisLeuIleGluSerPro 326
                       |||                           |||::::::   |||
```

```
Db    905 TTGATTACTCTTCCT--------------------ACTGCCCATGTGATGCCGTCTAAT 943
Qy    327 SerLysGlyArgThrThrLys-----------------------------------Ser 334
Db    944 TCCAGCGCTTCAATTTCCAAACTTAGGGAATCATTGACACCAGATGGCTCAAAATGGAGT 1003
Qy    335 HisIleAspAsnGlnProLeuSerSerSerSerGlnGlyGluThrSerPheSerThrTyr 354
Db   1004 ACAAGCCTCATGCAAACATTGGGAAACCACGGTAGGGGGGAGCAGGACTCTTCCCTAGAC 1063
Qy    355 TyrGluSerPhePro---------------SerSerGlyGlyGluGlyAlaIleProGlu 369
Db   1064 ATGAAGGACTTCCGGCCACTTCGGAAATGGTCATCTTTATCCAAACTCACTGCCCCGGAT 1123
Qy    370 ProSerArgSerAsnGlyLeuAlaArgSerGluGluSerAlaArgSerGlnValGlnVal 389
Db   1124 AACTGTGGCCAGGGTGGCACTGTATGCAGAGAAGAGTCAAGGAATGGTTTGGAGAAGATA 1183
Qy    390 ---HisAlaProValValAlaAlaArgLeuArgAsnIleTrpProLysPheProLysTrp 408
Db   1184 GGGAAGGCTAAGGCTTTAACATCACAACTGAGGACAATTGGG---------CCTAGCTGT 1234
Qy    409 LeuHisGluAlaProLeuAlaValAlaTrpGluValThrArgLeuPheMetHisCysLys 428
Db   1235 TTACATGAT---------------AGTATGGAGATGCTTAGG------------------ 1261
Qy    429 ValAspLeuGluAspGluSerLeuGlyLeuLysTyrAspProSerTrpSerThrAlaArg 448
Db   1262 ------TTAGAGGACAAGGAAATAAATAAAAAAACGG------------TCATCAACTCTG 1303
Qy    449 AspValThrAspIleTrpLysThrLeuTyrArgLeuAspAlaPheArgGlyLysProPhe 468
Db   1304 GACTGCAAG------------------TATAAA-------------------------- 1318
Qy    469 ProGluLysProProAsnAspValPheValThrAlaMetThrGlyAsnPheGluSerLys 488
Db   1319 -----------------------TTTGAGAGCTGTAGCAAGGAGGACTTTAGAGCCTCT 1354
Qy    489 GlySerAlaValValLeuSerAlaValLeuAspTyrAsnProAspAsnSerProThrAla 508.
Db   1355 TCCTCTACTCTTAGGAGACAGACTGTA---------------GACATGACATATAGTGCC 1399
Qy    509 ProLeuTyrLeuValLysLeuLysProLeuMetPheGluGlnGlyCysArgLeuThrArg 528
Db   1400 ---------TTACCTGAAAGCAAGCCCATTATGACAAGC---------------TCAGAG 1435
Qy    529 ArgPheGlyProAspArgPhePheGluIleLeuIleProSerProThrSerThrSerPro 548
Db   1436 GCTTTTGAACCTCCAAAATAT------TTAATGCTTGGTCAACAGGCAGTAGGTGGAGTT 1489
Qy    549 SerValProProValValSerLysGlnProGlyAlaValGluGluValIleGlnTrpLeu 568
Db   1490 CCCATTCAGCCTTCCGTAAGGACTCAG-----------------------ATGTGGCTT 1525
Qy    569 Thr------MetGlyGlnHisSerLeuValGlyArgGlnTrpArgAlaPhePheAlaLys 586
Db   1526 ACAGAGCAGCTGCGGACAAATCCTTTGGAAGGTAGAAATACAGAGGATTCTTACAGTTTA 1585
Qy    587 AspAlaGlyTyrArgLysProLeuArgGluPheGlnLeuArgAlaGluAspProLysPro 606
Db   1586 GCTCCTTGGCAACAGCAGCAAATTGAAGACTTTCGACAAGGAAGTGAAACACCAATGCAG 1645
Qy    607 IleIleLysGluArgValHisPhePheAlaGluThrGlyIle---------ThrPheArg 623
Db   1646 GTTTTG--------------------------ACTGGATCATCTCGTCAAAGTTATTCA 1678
Qy    624 Pro---------AspValPheLysThrArgSerValValProAlaGluGluProValGlu 640
Db   1679 CCTGGCTATCAGGATTTCAGTAAGTGGGAATCAATGTTGAAAATAAAAGAAGGACTTCTA 1738
Qy    641 GlnArgThrGluPheLysValSerGlnMetLeuAspTrpLeuLeuGlnLeu--------- 657
Db   1739 AGGCAGAAAGAAATTGTAATCGATCGGCAGAAGCAACAAATTACCCACCTGCATGAGAGG 1798

Qy    658 ------AspAsnAsnThrTrpGlnProHisLeuLysLeuPheSerArgIleGlnLeuGly 675
Db   1799 ATAAGGGATAATGAATTACGGGCTCAACATGCCATGTTAGGACATTATGTA------AAT 1852
Qy    676 LeuSerLysThrTyrAlaIleMetThrLeuGluProHisGlnIleArgHisHisLysThr 695
Db   1853 TGTGAGGATTCTTAT---GTGGCTAGTTTGCAGCCA---------CAATATGAGAACACT 1900
Qy    696 AspLeuLeuSerProSerGlyThrGlyGluValMetAsnAspGlyValGlyArgMetSer 715
Db   1901 TCACTCCAGACACCATTTTCAGAGGAATCAGTTTCCCATTCCCAACAAGGGGAATTTGAG 1960
Qy    716 ArgSerValAlaLysArgIleArgAspValLeuGlyLeuGlyAspValProSerAlaVal 735
Db   1961 CAAAAGCTTGCATCTACTGAGAAAGAAGTTTTACAGCTTAATGAGTTTCTCAAACAAAGA 2020
Qy    736 GlnGlyArgPheGlySerAlaLysGlyMetTrpValIleAspValAspAspThrGlyAsp 755
Db   2021 CTAAGCCTATTTAGTGAAGAGAAGAAGAAACTA-------------------------- 2053
Qy    756 GluAspTrpIleGluThr---------TyrProSerGlnArgLysTrpGluCysAspPhe 772
Db   2054 GAGGAAAAACTTAAAACTAGAGATCGATACATCAGTAGTCTGAAAAAGAAATGCCAGAAG 2113
Qy    773 ValAspLysHis------------GlnArgThrLeuGluValArgSer------------ 784
Db   2114 GAATCTGAACAAAACAAAGAAAAGCAGAGAAGAATTGAGACCTTGGAAAAGTATCTGGCT 2173
Qy    785 ---------ValAlaSerGluLeuLysSerAlaGlyLeuAsnLeuGlnLeuLeuProVal 801
Db   2174 GATCTTCCAACTCTAGATGATGTACAGAGTCAGAGTCTACAGCTGCAGATTCTGGAAGAA 2233
Qy    802 LeuGlu---------------------AspArgAlaArgAspLysValLysMetArgGln 814
Db   2234 AAAAAATAAGAATTTACAAGAGGCTTTGATAGATACAGAAAAAAAACTTGAAGAGAT--- 2290
Qy    815 AlaIleGlyAspArgLeuIleAsnAspLeuGlnArgGlnPheSerGluGlnLysHisAla 834
Db   2291 -----------------------------CAAAAAGCAGTGTCAAGATAAAGAGACACA 2320
Qy    835 LeuAsnArgProValGluPheArgGlnTrpValTyrGluSerTyrSerSerArgAlaThr 854
Db   2321 GTTAATATGCCAGAAAAAGAAAGAAAAGGAGTTAGTAACTACCGTTCAGAGTTTGCAACA 2380
Qy    855 Arg-----------------------------ValSerHisGlyArg-ValProPheLe 864
Db   2381 AAAAGTAGAAAGATGCCTTGAAGATGGAATCCGCCTTCCCATGTTAGATGCAAAACAGCT 2440
Qy    864 uAlaGlyLeuProAspSer---------GlnGluGluThrLeuAsnPheLeuMetAsnSe 881
Db   2441 TCAGAATGAAAATGATAATCTCAGACAACAAAATGAGACTGCTAGTAAGATAATAGACAG 2500
Qy    881 rGlyPheAspProLysLysGlnLysTyrLeu------GlnAspIleAlaTrpAspLeuGl 899
Db   2501 CCAACAAGATGAGATTGACAGAATGATTTTAGAAATTCAGTCTATGCAAGGAAAGCTTTC 2560
Qy    899 nLysArgLysCysAspThrLeuLysSerLysLeuAsnIleArgValGlyArgSerAlaTy 919
Db   2561 TAAAGAGAAACTGACCACTCAAAAG---------------------------------- 2585
Qy    919 rIleTyrMetIleAlaAspPheTrpGlyValLeuGluGluAsnGluValHisValGlyPh 939
Db   2586 -------ATGATGGAAGAG------------CTGGAAAAGAAAGAAAGAAATGTACAGAG 2626
Qy    939 eSerSerLysPhe---------------ArgAspGluGluGluSerPheThrLeuLeu-- 953
Db   2627 ATTAACAAAAGCATTGCTTGAAAACCAAAGACAAACAGATGAGACATGCTCTTTATTGGA 2686
Qy    954 ---------------------SerAspCysAspValLeuValAlaArgSerProAlaHi 966
Db   2687 TCAGGGCCAGGAACCTGACCAATCTAGGCAGCAGACAGTTCTTTCCAAACGGCCACTATT 2746
```

```
Qy        966 sPheProSerAspIleGlnArgValArgAlaValPheLysPro----------------- 980
                        |||:::      :::      :::||||||
Db       2747 T---------GATTTGACTGTGATTGATCAGCTGTTCAAGGAAATGTCCTGTTGTTTGTT 2797

Qy        981 -GluLeuHisSerLeuLysAspValIleIlePheSerThrLysGlyAspValProLeuAl 1000
                :::|||   :::|||      ::::::            :::|||      |||
Db       2798 TGACTTGAAAGCATTGTGTAGTATTCTTAATCAGCGTGCTCAGGGCAAGGAGCCTAATCT 2857

Qy       1000 aLysLysLeuSerGly 1005
                     |||   |||
Db       2858 TTCATTATTACTGGGA 2873
```

Search completed: March 31, 2004, 12:59:51
Job time : 5497.3 secs